Master Thesis in Geosciences

Modelling of rockfall runout range

Employing empirical and dynamical methods

Martin Weme Nilsen

UNIVERSITY OF OSLO

FACULTY OF MATHEMATICS AND NATURAL SCIENCES

Modelling of rockfall runout range

Employing empirical and dynamical methods

Martin Weme Nilsen

Master Thesis in Geosciences

Discipline: Environmental geology and geohazards

Department of Geosciences

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

[02.06.2008]

Tutor(s): Kaare Høeg (UIO/ NGI) and Ulrik Domaas (NGI)
This work is published digitally through DUO – Digitale Utgivelser ved UiO
http://www.duo.uio.no
It is also catalogued in BIBSYS (http://www.bibsys.no/english)

Cover photo: Photography taken from a rockfall event at Mundheim (Norway) (Domaas).

Abstract

Literature studies considering the determination of rockfall runout range have shown that the utilization of rigorous physical models yields significantly different results. This is primarily due to the fact that rockfall is a complex motion involving several modes of movement from free fall to bouncing, rolling and sliding. This has lead to difficulties in assigning the loss factor determined by the coefficient of restitution. Furthermore, investigations have shown that the restitution is dependent on many variables other than just material properties. The significant variance of variables and a limited knowledge of the dependence of each physical factor in relation with the retardation, make detailed numerical modeling of the phenomenon difficult and time consuming.

As an alternative we may utilize general terrain parameters common for all mountainsides to derive the runout distance. This thesis has focused on the statistical relations between the parameters, and how these can be used to derive an approximated runout range. In this respect two are based on simple linear regression results, while one is based on a multiple regression. In addition two simplified dynamic models have been put to the test. The results indicate that simple linear regression models may yield a relatively good first estimate of runout distance. However, the best statistical result was found with the use of a multiple regression model.

Keywords: rockfall, runout, dynamic modeling, empirical modelling, coefficient of restitution, dynamic friction coefficient, talus, topographical parameters, linear regression.

Table of Contents

1. Introduction

Rockfall is a slope process involving the detachment of rock fragments and their fall, followed by bouncing, rolling, sliding and deposition near the foot of the slope. It may occur on natural cliffs, excavated rock faces, or steep exposures of coarse-grained soils (Hungr 1989). The collective term of rockfall is signified to cases in which the volume is less than $100m^3$, as opposed to rockslides where the detached volume is set between 100-10000 m^3 (Domaas 1985).

The fall is initiated with the detachment of rocks from bedrock slopes, which is mostly a cliff face in the case of rockfall source area. The weathering process will then transpire along existing discontinuities such as joints, bedding planes, faults and cleavage surfaces. To a minor extent, development of new discontinuities (cracks) aids the process of detachment (Hungr 1989).

Although rockfall may be initiated at slopes covered with vegetation, the characteristic evidence left behind by rockfall activity is the talus slope. This is signified by scree deposits at the foot of steep cliff faces (Hungr 1989). In this respect an indication that the blocks descends from rockfall and not rockslides, is a graded deposition pattern with smaller blocks at the upper part of the talus and progressively larger blocks further down (Domaas 1985)

Rockfall is considered to be a relatively small landslide confined to the removal of individual and superficial rocks from a cliff face. Even so, it has the potential to generate large-scale mass movements of rock material. Such more extensive processes are defined as rockslides or rock avalanches. As opposed to fragmental rockfall, such stürzstroms are known to invariably disintegrate and proceed by flow like motion. In some rare cases rockfall has also been known to trigger catastrophic debris streams, which are even more hazardous. Because of its unpredictability in frequency and magnitude, rockfall poses a significant hazard for residential areas and infrastructure in mountainous areas (Dorren 2003).

The most common processes which cause rock fractures to open or propagate are frost action and water pressure increase due to rapid infiltration of snowmelt or precipitation. In addition to this, other types of physical weathering like crystal growth, hydration of minerals, root penetration, stress relief and frost-thaw processes may contribute to instability (Hungr 1989).

Currently there exist a number of mitigation methods. The most common being the constructing of catch or barrier fences and restraining nets. A more cost effective and sustainable method is the maintenance of forest stands or protection forests. In general the energy dissipated due to collisions with other rocks or obstacles in their trajectory is an important aspect of rockfall. Even so, these influences are difficult to analyze during rockfall events or in the field, and thus represents a challenge for future development (Dorren 2003).

1.1 Scope and background

This thesis intends to enlighten some of the questions surrounding the different types of rockfall modelling. Currently there exist a number of methods to estimate the runout range of rockfall events. All existing rockfall models can be categorized into three main groups: empirical (statistical) models, process based (dynamic) models and GIS-based models (Dorren 2003). This thesis will focus on the two former methods.

The sample based on this study is a set of 122 reported rockfalls with different parametrical variables collected from various parts of Norway. The survey was performed with a 50 m long measurement band and a simple elevation device. The range of the blocks beyond the talus was measured by utilizing the measurement band along the ground perpendicular to the contour line. The different topographical angles with respect to the block endpoint, cliff height and talus was measured with the elevation device (Domaas 1985).

In locations where difficult terrain conditions prevailed, the block was measured with respect to reference points such as roads or fences before being plotted to the map. All collected measurements was used to plot the block, the runout zone beyond the talus, the top of the talus and the top of the mountainside to a map in a scale of 1:5000 and equidistance of 5 meters. This was then transferred to a terrain profile (Domaas 1985). The collected parameters for each location can be found in the back of this report (see Appendix A).

1.2 Problem

Rockfall is a regular process in mountainsides and a significant hazard to infrastructure and homes in mountainous environments. It is evident that the most important way of reducing the risk of damage to an area is to have a comprehensive knowledge of the parameters and physical relations which govern the runout distance or runout zone. However, the modes of motion in rockfall involve several stages from freefall to toppling, rolling and sliding. The importance of these types of movement varies both temporally and spatially. Furthermore, the trajectory of the boulders is significantly sensitive to topography, as well as volume, size and shape. These factors make numerical modelling of the rockfall process very complicated. As an alternative approach we can utilize empirical relations based upon generalized topographic parameters of a mountainside to predict the runout of extreme runout distances (Keylock 1999).

1.3 Structure of thesis

The structure of the thesis is divided in two general parts, where each part is intended to describe or serve as a basis for a dynamic or empirical modeling approach.
Chapter one is a general introduction to the theme of rockfall. It presents the problem of rockfall modeling and the scope and background to the problem. The second chapter specifies some of the terrain parameters which are thought to be central within the model frame.

The third chapter serves as a basis to the general methodology utilized in dynamic modeling. Furthermore, some of the physical and geomorphologic aspects of the central part of the boulders movement area in the mountainside, that is the talus, are presented. Chapter four describes the mathematical framework and associated significant factors within general statistical operations, while chapter five outlines the methodology and mathematical basis of linear regression analysis.

The next two chapters intend to evaluate and discuss the results of the conducted analysis. In the sixth chapter the most important results of the dynamical analysis are presented and judged, while chapter seven evaluate the result of the empirically based models and the relations between some of the topographical parameters. The thesis ends with a general validation of the model approaches and a conclusion to the findings.

2. Topographical parameters

Empirical or statistical models are based on relationships between simple topographical parameters related to the rockfall slope and the length of the runout zone. For each rockfall case the geometry of the mountainside is described as in figure 2-1 or Appendix A, with release area as the uppermost cliff (H_3, θ), the talus itself with top and toe (H_2,Ψ), and the area beneath the toe of the scree (H_1,γ) (Domaas 1994). In view of the fact that rockfall is a complex motion, involving several modes of motion which varies both temporally and spatially, it is reasonable to derive different statistical relations based on simple topographic parameters (Keylock 1999).

2.1 General overview of topographical parameters

In order to make the calculations of empirical models feasible, we have to make some simplifications of the topographical parameters which we employ. These arrangements have been a necessity to compare the different cases. In addition, singular cases where particular topographic features dominate the result has been left out of the sample (Domaas 1994).

Figure 2-1 represents the generalized image of the different variables collected for each case. The mountainside is divided into three general parts according to the topographical angles of the rockfall hill. From the lower part and up to the scarp of the source area these are as follows:

- γ which signify the angle of the runout zone (S1) with respective height H1
- δ which signify the angle of the talus with respective height H2
- θ which signify the angle of the cliff face with respective height H3

The three potentials combine to form the total height (Htot) of the mountainside, while the horizontal distance can be divided in a similar manner with L_1, L_2 and L_3 connected to their respective height. The range is expressed by two angles: from the block to the talus peak (α), and from the block to the top of the cliff (β). Furthermore the area between the talus toe and the top of the mountainside is described by means of the ψ-angle. As a result we have six angles to provide a description for a given mountainside (Domaas 1994). Three of these six angles express the inclination on different parts of the mountainside (γ, δ, θ); while two describe the range of the rock with respect to different parts of the mountainside (α, β). Lastly, the ψ-angle describes the range of the talus toe with respect to the uppermost part of the scarp.

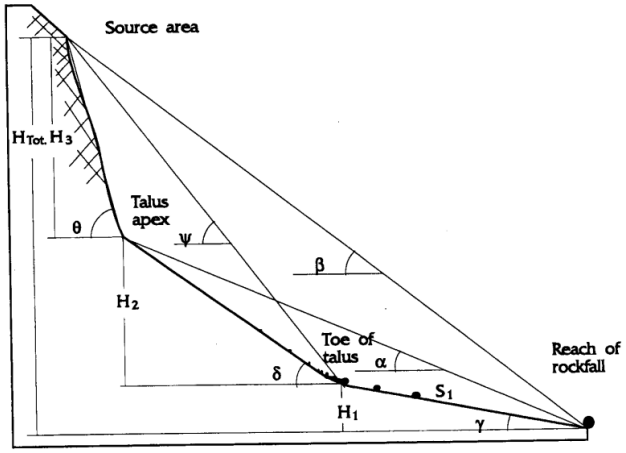
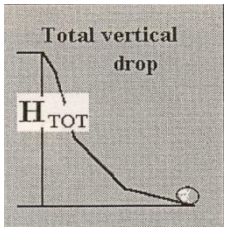


Figure 2-1:Topographic angles measured on slope profile (Domaas 1994).

2.1.1 *Parametrical variable study*

The angles are defined with respect to the inclination of the terrain, or with respect to the empirical angles of the mountainside. To ensure accessibility these are based on features which are detectable on all profiles (Domaas 1994). It has to be said that the simplification to that of general empirical parameters is a significant simplification. Even so, it is beneficial in the way that relevant parameters can be compared.

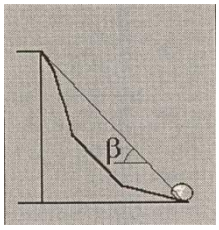
2.1.2 *Total height of the cliff face*



The runout of falling blocks for long distances is regularly presumed to depend on the height of the slope. This is a fair judgment since the velocity of the rock increases with the potential during free fall. The total height (htot) of the mountainside is defined as the total vertical difference between the top scarp of the source area and the block location. The runout (S_1 on figure 2-1) has been estimated with a tape from the talus toe, while the base of the talus is expressed as the lower boundary of the area which is entirely covered

with debris. Difference of opinion with respect to the talus base is implausible to influence the result with more than a few meters (Domaas 1994).

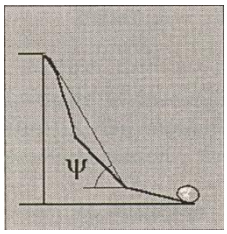
2.1.3 *The β-angle (Fahrböschung-angle)*



The β angle may be described as an equivalent to Albert Heim's fahrböschung angle. It describes the maximum runout of a body for a given mountainside and is defined as an angle from the block to the top of the uppermost scarp. The angle is used by Scheidegger as a basis for a description of the extension of an average friction coefficient for landslides. In general β=30 has been used to give an estimate of the run out distance for rockfalls. The factor is dependent of the height of the mountainsides and the runout distance of the boulder.

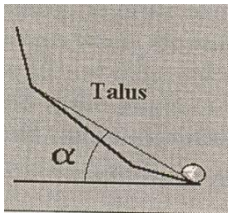
Usually β-values are greater than 31^{o} for mountainsides with heights between 100 and 350 meters. From then on the values increase linearly to 35^{o} for a potential of 650 m (Domaas 1994).

2.1.4 *The ψ-angle*



In some cases engineers are faced with the question of how far rockfall can reach in developed areas where earlier rockfalls has been cleared. The potential of the mountainside and the end of the talus can be used as a basis for the determination of this. The ψ-angle is defined by an angle going from the talus toe to the uppermost scarp of the mountainside. Even though increasing heights would be assumed to increase the run out, it is important to note that the loss of energy due to impact also increases with the height. Because of this effect the runout distance would be smaller than the relative difference in potential would seem to indicate. The ψ-angle describes the steepness of the mountainside irrespective of the total height (Domaas 1994) .

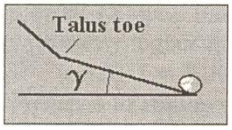
2.1.5 *The α-angle (shadow angle)*



Formerly an α-value of 25^{o} has been used as a general value for calculation of how far rocks may travel outside the toe of the talus. This value is however too inaccurate to utilize for practical calculations. Measurements from former investigations indicate that the angle is geometrically dependent on the height and length of the talus. Subsequently one of these parameters should be applied together with α to describe rockfall range. The talus is the result of local topography and former rockfalls which have built the

talus since the ice age. The depositions of the falling blocks are constituted by the topography of the mountainside and the transition to the terrain below. Because of this inconsistency the talus peak does not provide a reliable reference for investigation of run out (Domaas 1994).

2.1.6 *The γ- angle*



The γ-angle denotes the distal inclination of the mountainside. The corresponding length beyond the talus is expressed by the runout (S1,fig.2-1, Appendix A). This part of the mountainside is referred to as the talus shadow by Hungr. Since it is the most distant and atypical observation it ultimately represents the most hazardous part. In former reports by Domaas it is stated that the longest runouts are related to the slope γ, and that the γ-angle has a particularly great effect when the terrain beyond the talus toe is steeper than 15^{o} (Domaas 1994).

3. Methodology in dynamic models

The dynamic models use physical equations to estimate the runout of a rockfall. These types of models describe or simulate the modes of motion of falling rocks over a slope. The models are simplified and do not take the fragmentation of rocks upon impact or interaction with other rocks into consideration (Larsen 1993). As a result of not taking this into account, they are based on a conservative approach (Azzoni 1993). This is because a constant size and mass which does not break apart after impact, ultimately represents the worst case scenario (Pfeiffer 1990).

Although there is a large group of dynamical models they all have some common factors. First, they are all two-dimensional slope-scale models which restrict boulders to move in a vertical plane. Thus, lateral movements are not simulated. Secondly the track on which the boulder follows is defined as a set of lines with a slope angle equal to the measured mean slope gradient. Lastly the movement is simulated as a succession of flying phases and contact phases (Dorren 2003).

3.1 Physical relations in dynamic models

The motion of rockfall usually involves a rock falling from a cliff, followed by bouncing rolling and sliding, before the frictional retardation eventually brings it to a halt. Even if we ignore the large random component in this motion, it is evident that this is a complex motion to describe mathematically.

As an alternative, Statham and Kirkby outlined a first order model where the average movement can be treated as though the rock is sliding on a rough surface.

In effect this makes the model more feasible since the elastic responses of the rock and talus are simplified, while non-elastic effects like shattering and crating are ignored. The interaction between the rock and the surface is described by the dynamic angle of friction $\varphi'_{\mu d}$, which experiments has shown to be a constant (Statham 1975).

The general scheme of Statham and Kirkby's frictional model initiates with a cliff of a height h, and a rock descending towards the talus slope at an angle δ. By applying the principle of free fall energy conservation and neglecting air resistance, the falling rock will have the following velocity at the base of the rock (Statham 1975):

$$v = \sqrt{(2gh)} \qquad (3.1)$$

Where (v) is the velocity in m/s, h is the height of the cliff and g is the gravitational acceleration equal to 9.81m/s^2. At impact with the talus, the component of the free fall velocity is conserved, thus producing an initial down slope velocity (u) of (Statham 1975):

$$u = \sqrt{(2gh)} \sin \delta \quad (3.2)$$

While at the scree surface the rock is subjected to a downslope component of its weight which is equal to$mg \sin \beta$. Conversely, the force of resistance contributed by the frictional force working against the motion follows the theory of Coulomb's law of dry friction (Bozzolo 1986):

$$\mathrm{F}_f = \mu_f \cdot N = \mu_f \cdot mg \cdot \cos \delta \qquad (3.3)$$

Where μ_f is equal to the tangent of the dynamic friction coefficient $(\phi\mu'_d)$. Subsequently, in a given segment where the frictional force of resistance is greater than the driving force, the rock will experience a deceleration (a_δ) equal to (Statham 1975):

$$a_\delta = g\left(\sin \delta - \cos \delta \tan \phi'_{\mu d}\right) \qquad (3.4)$$

Following this outline the rock will come to rest at a mean distance ($\bar{x}$) equal to (Statham 1975):

$$\bar{x} = \frac{u^2}{2\alpha_\delta} = \frac{h \cdot sin^2 \delta \, cos \, \phi'_{\mu d}}{sin(\, \phi'_{\mu d} - \delta)} \qquad (3.5)$$

This expression may be rearranged to give an expression for $\phi'_{\mu d}$ (Statham 1975):

$$\tan \phi'_{\mu d} = \left(1 + \frac{h}{\bar{x}} \sin \delta\right) \cdot \tan \delta \qquad (3.6)$$

Thus, (3.5) and (3.6) implies that as h/x tends to zero, the slope angle δ approaches $\phi'_{\mu d}$. Furthermore, we see that lower values of $\boldsymbol{\phi}'_{\mu d}$ will in effect yield greater values of mean runout distances $(\bar{x})$, given a constant fall height, and that distance of travel should be proportional to the cliff height (H3) (Statham 1975).

Alternatively, the velocity for a rolling rock on a path segment of a constant slope may be written as a general comprised equation (Hungr 1989):

$$v = \sqrt{v_0 + 2gx(\sin \delta - \mu_f \cos \delta)} \qquad (3.7)$$

Where v_0 is the acquired velocity (m/s), x is the rolling path distance of the point mass (m), while δ is the mean slope gradient in degrees. By employing this calculation, the rolling friction force will be the most determining factor for the velocity (Hungr 1989). Thus the energy dissipation will only be due to the resistance of rolling (Domaas 1994).

It is important to note that the model is simplified and will only model mean distances of runout. Additionally it implies neglect of energy absorption by the scree surface. In reality an increased absorption effect will occur with the difference in size between significantly large rocks and comparatively smaller stones on the scree surface (Statham 1975).

According to Staham and Kirkby this effect becomes substantial when the ratio of sizes exceeds 1:10. Furthermore, it is recognized that small rocks that detach from large cliffs may reach their terminal velocity in air prior to impact with the bottom of the cliff (Statham 1975).

3.2 Coefficient of restitution

The coefficient of restitution expresses the amount of energy dissipated during impact on the ground. Founded on Newton's theory of rigid particle collision, Szabo states that particle restitution is not merely a material constant, but is also dependent on velocities (rotational, translational), material type and geometry. In context with rockfall studies the coefficients of restitution are often considered to only be a function of the slope material. This assumption implies that they are independent of configuration, direction of approach, velocity and friction. Thus, they are based on a rough description of the slope material, occasionally complemented by data according to the roughness, compaction degree and vegetation cover (Heidenreich 2004).

Even so, former studies of collisions with the substrate show that the movement of the block at the impact end is affected not only by features of the slope, but also by several parameters related to the boulder as well as to the kinematics during impact. Figure 3-1 states some of the most important parameters believed to influence the restitution.

Slope characteristics	Block characteristics	Kinematics
• strength • stiffness • roughness compared to block size • inclination	• strength • stiffness • weight • size • shape	• velocity (translational and rotational) • collision angle • configuration of the block at impact

Figure 3-1: Table of parameters assumed to affect the bouncing of the object (Heidenreich 2004).

Because of the interaction of these factors, rockfalls originating from the same source may display different behaviour (Pfeiffer 1990). In order to ensure even and repeatable conditions, the majority of the parameter studies have been performed in laboratories at small scale.

The parameter may by estimated by applying two different methods through in situ tests. These are the following (Azzoni 1993):

- evaluation of the ratio between the energy after and before the impact.
- back analysis of previous falls.

Back analysis considers the tracks left in the field or on trees and the observation of block desposits. This may help in understanding the different aspects of the phenomenon such as range of block size, horizontal as well as lateral runout distances, bouncing heights and length (Heidenreich 2004).

Rigid body models usually utilize only one restitution coefficient generally based on the energy ratio, while other models usually adopt normal and tangential restitution coefficients. These are based on the translational velocity before and after the impact and indicated as K_n and K_t for normal and tangential restitution coefficients respectively (Azzoni 1993).

In a conservative worst case perspective, the falling body is often considered to have the shape of a sphere or cylinder. The following impact equations for tangential (3.8) and normal (3.9) velocity components respectively, may then be applied in a rigouros model.

$$v_t = \frac{5}{7} v_i \sin\theta - \frac{2}{7} R\omega_i \qquad (3.8)$$

$$v_n = -K_n v_i \cos\theta \qquad (3.9)$$

Where v_i is the incident velocity, while θ corresponds to the angle with the normal to the surface, ω_i denotes the incident angular velocity and R is the sphere radius (see fig.3-2) (Hungr 1989). Accordingly, a particle which has no rotation prior to collision with the substrate will experience a reduction in its tangential momentum by a factor of 5/7 (71%), and commence rotation after impact with an angular velocity 71 per cent of the incident tangential velocity.

In general it has been found that compact rock slopes have higher restitution coefficients than debris or earth slopes (Heidenreich 2004). However, the variation within same materials are significant.

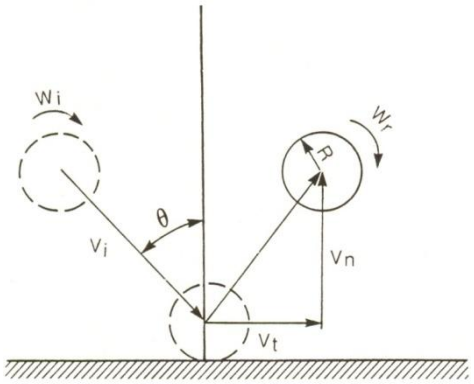


Figure 3-2: Illustration of an ideal collision of a spherical body against a plane surface (Hungr 1989).

3.3 Modes of motion

After a rock has detached from a cliff and starts to move, it descends from a cliff in different modes of motion. The modes of motion after detachment of the rock are dependent on the slope geometry. The three most important modes of motion are free fall, bouncing on the slope surface and rolling over the slope surface.

Freefall occurs on steep slopes with a transition from bouncing above 70^o. In this mode of motion we have two types of motion; translation of the centre of the rock or rotation of the rock around its centre. The air friction has no significant effect on the rockfall and may therefore be considered negligible (Dorren 2003).

About 75-85% of the energy is lost during the first collision with the slope. In addition, the initial bounce will also be likely to break the rock. Thus, much of the energy gained from a long free fall is lost in the first impact with the talus (Hungr 1989). That being said, the relative difference in the acquired velocity makes the height of the cliff a significant contributor the length of the runout.

If the slope gradient decreases to under 45^o a bouncing rock will gradually transform its motion to rolling because the rock gathers rotational momentum. The transition from

bouncing to rolling movement will occur much sooner in case of large blocks as opposed to smaller blocks (Hungr 1989). In this respect the shape is also of significance. This is because blocks of round shape achieve longer run out distances (Petje 2005).

While changing movement from bouncing to rolling, only the edges with the largest radius maintain contact with the slope. In effect the centre of gravity moves along an almost straight path, which is an effective mode of motion with respect to energy loss. Sliding motion will generally occur only in the initial and final stages of a rockfall (Dorren 2003). The sliding process is governed by the frictional properties of the particle to surface contact, and when it is about to occur we have the following relation (Statham 1972):

$$\mu = \frac{w \sin \alpha}{w \cos \alpha} = \tan \alpha = \tan \varphi_{\mu} \quad (3.10)$$

Where μ is the coefficient of friction, w is the weight of the block, α is the slope angle and φ_{μ} is the sliding friction angle (Statham 1972).

Eventually after experiencing different modes of motion the rock will stop. The velocity and subsequently halt of the rock is primarily dependent on the mean slope gradient. The reason for this is that falling rocks often accelerate on steeper slopes and decelerate on flatter slopes. In addition to the mean slope gradient the velocity is also dependent on the size of the rock and on the material covering the slope (e.g. soil, scree and vegetation). The latter is related to the retardation in connection with the coefficient of restitution and roughness of the slope. Small rocks will retard more easily than larger rocks. The reason for this is threefold; first the total kinetic energy of a small rocks is comparatively lower than for larger rocks, secondly large obstacles such as trees can more easily stop small rocks and lastly small rocks retard more easily in depressions between larger rocks in talus slopes. These are the main reasons why the fine material is often found near the base of the talus slope, while the larger blocks are found further downslope (Dorren 2003).

3.4 Lumped and rigorous models

In what is referred to as a lumped mass modelling perspective, the mass of the falling rock is concentrated in one point The point is then considered to follow a ballistic trajectory in the air, while neglecting air resistance (Dorren 2003).

At impact the velocity vector components, normal (V_n) and tangential (V_t) to the slope surface, are reduced in accordance with the coefficients of restitution (K_n, K_t) in the following manner (Hungr 1989):

$V_{nred.} = V_n K_n \quad (3.11)$
$V_{tred.} = V_t K_t \quad (3.12)$

One significant simplification that these models use is that the block (modelled as a point) has only translational velocity and no rotational velocity (Azzoni 1993).
Instead the two coefficient values are selected as bulk measures of impact characteristics like translational momentum, deformation and contact sliding. Consequently they are dependent on fragment shape, slope, surface roughness, momentum and deformational properties (Hungr O, 1989). An alternate method is to use a coefficient of defined energy loss at impact, defined by the ratio of total kinetic energy before and after the impact (Azzoni 1993, Alejano 2007).

Rigorous models incorporate the shape and volume of the rock (Alejano 2007).
One such model developed by Falchetta considers the rock as a rigid and non deformable body, while the slope is a sequence of segments with determinate length and mechanical characteristics. While based on fundamental laws of physics, the model is very complex since every reaction force at contact with the slope is dependent on the particular conditions of the impact. Furthermore, the generated force is defined only when its point of application is known. Subsequently a procedure based on time increments is required where the position of the block and the shape of both the block and the slope are assessed (Azzoni 1993).

The weakness of rigorous models lies in the strict deterministic approach which requires input of realistic data for many parameters. In addition the general complexity and long calculation time renders the practical use of many of these models. Nevertheless the accuracy of rigorous models is considered to be better than for lumped models (Azzoni 1993).

3.5 Significance of talus slopes

The internal structure of the talus is characterized by the material found between the exposed scree surface and the underlying face where talus accumulation has taken place. In this respect the underlying face is commonly represented by bedrock (Brown 1994).

Apart from talus accumulation models tested by Brown, and former short references to experimental talus modelling by Kirkby and Staham in 1975, the subsurface of talus deposits has not been subject to extensive research (Brown 1994). Even so, the topography and features of this part of the mountainside is of vital importance to the restitution of the boulder and its associated mechanical behaviour. Furthermore, in view of the fact that it is a prominent feature in rockfall dominated mountainsides, the internal structure and composition is crucial for an even better understanding of rockfall behaviour.

3.5.1 *Evaluation of talus subsurface*

There are a number of methods used to characterize a talus in literature. In this respect the most common are identification based on profile shape, slope angles particle size distribution, particle age distribution, particle size range and slope length.

There is a wide consensus among authors that the talus profile displays a general concave profile, or varying degree of concavity (fig. 3-3). In support of this, Statham concluded in his study of talus profiles at two sites that all slopes consist of a rectilinear upper slope and some degree of basal concavity.

The slope angles are primarily derived from the mean unconstrained slope angle which is defined as the mean angle of the slope, excluding the parts of convexity and the basal concavity. In a fractioned perspective, an angle of a slope segment is referred to as a facet angle. Studies of literature sources has shown that the unconstrained slope angle has a well defined range between about 30 and 38^{o} degrees, while the facet angles has a greater variance posing values between 10^{o} and about 40^{o}.

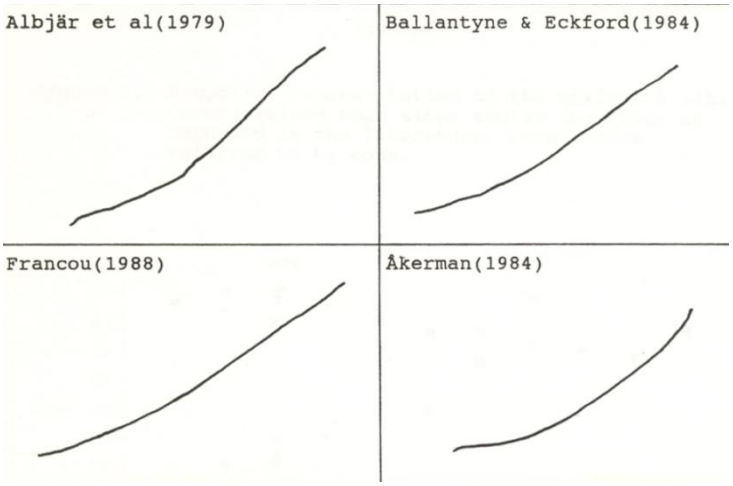


Figure 3-3: Illustration of examples of the talus profiles in literature (Brown 1994).

The particle size distribution describes how the particle sizes change along the talus slope. On a large proportion of talus there is a reduction of particle size upslope. This sorting is believed to be due to rockfall activity; hence it is often referred to as "fall sorting". Other processes such as avalanche and flow processes can modify this pattern, even to its removal in particular high activity cases.

According to a study by Bones in 1973, 10 out of the 15 rockfall talus slopes studied show a downslope sorting pattern. This result shows that fall sorting is an important part of talus characteristics.

In an investigation by Gardener it was reported that all slopes showed a logarithmic decrease in particle size upslope. Thus, the results suggests that the rate of change in size with distance upslope is not constant, varying from slope to slope.

It was found that shorter slopes showed a faster rate of change in grain size, while longer slopes displayed changing grain sizes in the lower parts which turned in to constant grain sizes in the upper parts.

Two investigations by Statham in 1973 and 1976 gave evidence of a linear relationship between slope position and grain size reduction. The reason for this was assumed to be due to the character of the mix of grain sizes supplied from the rockwall (Brown 1994).

3.5.2 *Talus stratum*

In an investigation by Rapp (1960) it is reported that talus material usually lies as a relatively thin layer of some few meters thickness over a substratum of bedrock with an inclination about the same as that of the talus surface. This assertion is confirmed by reports of maximum values of talus thicknesses (1-3m, 2-5m, 2-8m, up to 25m, and up to 35m), and by the results of his own investigations of 6 talus, giving maximum thicknesses of 8m, 18m, 20m, 22m, and 24m (Brown 1994).

This variance has also been found by Ballantyne and Eckford (1984) at a study of two different sites in Scotland. The uppermost layer in the talus material has been reported to a depth of 0.8 m according to Rapp (1960) and 1 m according to Åkerman(1984). They both report a net reduction in particle size with depth (Brown 1994).

According to Rapp the grain size range is held constant, but the proportion of sands and gravels increases with depth. Åkerman reports a gradual transition from coarser to finer material with depth. According to their findings there is a net reduction in particle size with depth.

Rapp finds that the range of grain sizes remains constant, but the proportion of sands and gravels increase with depth, while Åkerman reports of a gradual transition from coarser to finer material with increased depth.

Figure 3-4 displays the talus cross section according to Church et. al. It is important to note that it is not clear from the context of his work, what source material the figure originates from. Subsequently it is not clear if this is a figure based on a general summation of talus observations, or if it is a presentation of a single observed talus (Brown 1994).

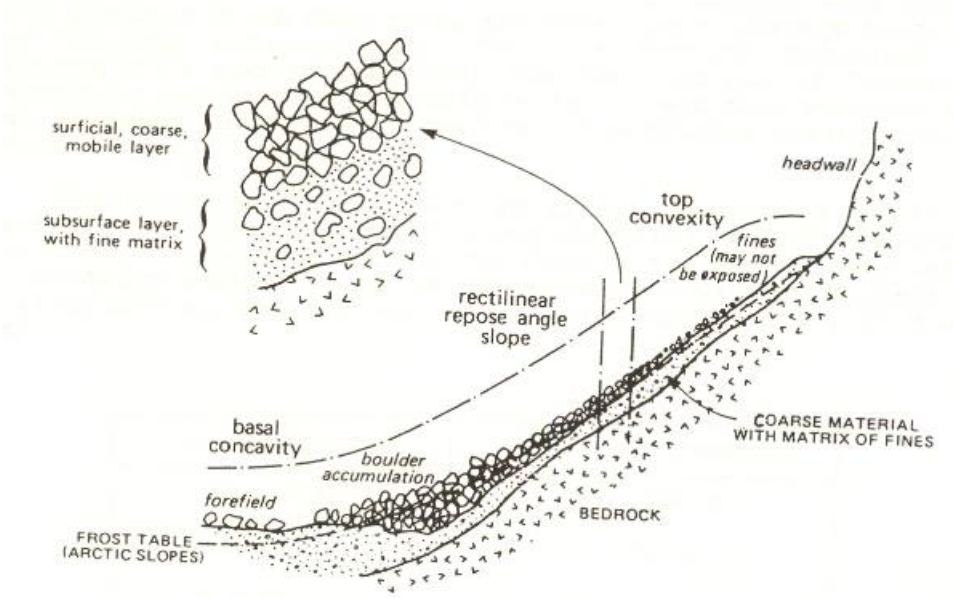


Figure 3-4: Figure displaying a cross section of a talus deposit as presented by Church et al. (Brown 1994).

3.5.3 *Movement on talus and influence of profile form*

A body falling on a scree surface will contribute to an energy proportional to the height of its fall from the mountainside. When the body impacts with the surface, the energy it carries is resolved into two components. These are respectively down-slope and normal to the slope.

During impact the talus will absorb the energy of the normal component and may tend to move other particles in the process. Meanwhile the body will commence its movement down slope with a velocity which is equal to the down slope component of the body's falling velocity. This velocity is proportional to the sine of the slope angle. The runout distance will increase with the average input of energy in the body. This is because a higher average energy implies a greater down slope velocity, which in turn makes the body able to travel greater distances from the headwall before stopping under the influence of friction.

In cases where the talus is short the particles will not come to a halt, but will move on to ultimately accumulate as a tailing concavity. That being said, if the average input of particles is small, or the talus slope long, the majority will come to an end at the base and the concavity will only occupy a small part of the total profile.

In the early stages of talus accumulation at the foot of a cliff, the mountainside is high. In effect the input of energy is high as well. As a consequence as soon as some accumulation has occurred, particles impacting on the slope obtain large down slope velocities.
Because of this, a large amount of the particles accumulate as a long tailing concavity which extends beyond the mountainside.

The height of the exposed mountainside becomes less with the accumulation of scree, leading to progressively smaller inputs of energy. At the same time, the increase in length of the talus allows only the high energy particles to travel as far as the concave section. This results in an increasing slope angle in the upper section that eventually reaches an equilibrium at which the addition of particles from the cliff is balanced by removal due to impact.

This tends to balance the net deposition along the slope length. Therefore the lower concave section declines in size and significance as the height of the cliff becomes less, and is replaced by a straight slope which is in equilibrium with the average input of energy possessed by the particles (Statham 1972). When the cliff height goes toward zero, the energy input is so small that the particles have nearly no down slope velocity. At this state the buildup is confined to the top of the slope until the maximum angle of stability for the loose state is reached. After this, dry avalanching of the granular material may occur.

Thus when reaching a steady state the headwall height approaches zero and the landform is a straight slope at the angle of repose. Until this state is reached the slope is subject to a progressive development related to the kinetic energy of the supply. Thereafter the principal activity ceases, because of the disappearance of the supply zone, and slope degradation must proceed by other processes (Brown 1994).

3.5.4 *Significance of talus backwall shape*

The inclination of the backwall, or cliff height, of a mountainside represents an important factor in talus accumulation and subsequent profile. Former field studies conducted by Kotarba (1988) divides the talus shape according to the inclination of the backwall (Brown 1994).

In Kotarba's type I accumulation the backwall is near vertical. This will in effect enable a rock particle to fall without coming in contact with the backwall. As a consequence, rotation will not be applied to the rock during the descent towards the talus. The resulting accumulation pattern will tend to form a steep straight talus deposit of parallel inclined layers (see fig.3-5) (Brown 1994).

Type III accumulation evolves under a more inclined (sub vertical) backwall. This will increase the potential for particle contact under the fall and a subsequent application of rotational velocity. The associated accumulation pattern will form a distinct concave profile, while the basal concavity section is described as being a relict inactive part of the talus (see fig.3-6).

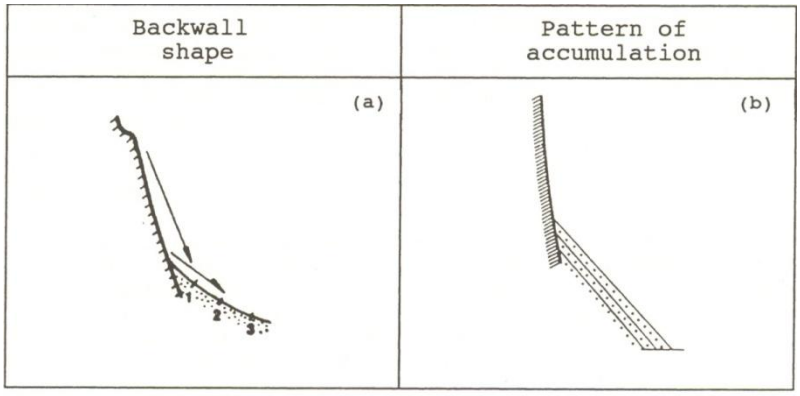


Figure 3-5: Illustration of Kotarba's type I accumulation pattern (Brown 1994).

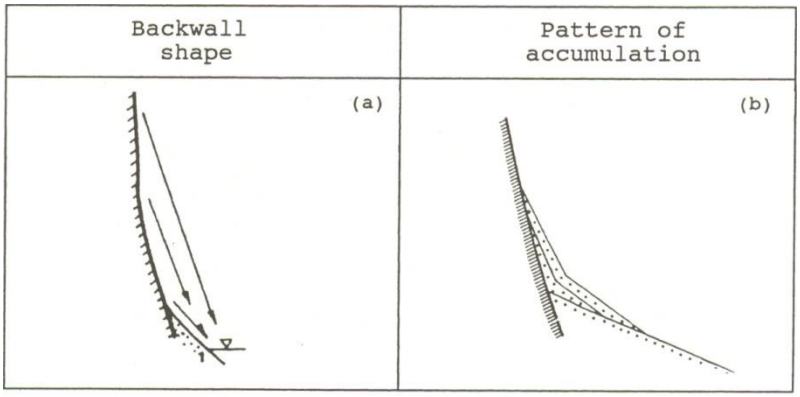


Figure 3-6: Illustration of Kotarba's type III accumulation (Brown 1994).

3.5.5 *The shadow angle's relation to the talus slope*

The rockfall shadow angle is determined as the angle between the end point of the boulder and the top of the talus slope. Furthermore, the rockfall shadow is an expression used to define the distal part of the slope which is discontinuously scattered by large boulders that have moved beyond the talus. Observations and analysis by Lied on natural rockfalls yielded angles in the interval 28-30^{o}. These observations are confirmed by Evans and Hungr, who stated minimum shadow angles of 27.5 degrees (Heidenreich 2004).

Empirically determined shadow limits may be an useful device to approximate maximum runout at the base of talus slopes. A rockfall dominated talus slope has a general profile as seen in fig.3-7. Commonly, the finest fragments accumulate below the apex (point A), at an approximate angle of 38^{o}. Further down the inclination of the talus ranges from approximately 32 to 38^{o}. The lower part of the talus holds the largest rocks sizes. In this section the slope occasionally falls between 10 and 20^{o}, as can be seen between points B and C.

The latter point is the base of the talus deposit and beyond this point the slope is no longer fully covered by talus particles. The average talus slope angle is stated as; min shadow angle (refer to fig.3-7), while the rockfall shadow angle is stated as; max shadow angle (Hungr 1989, Heidenreich 2004).

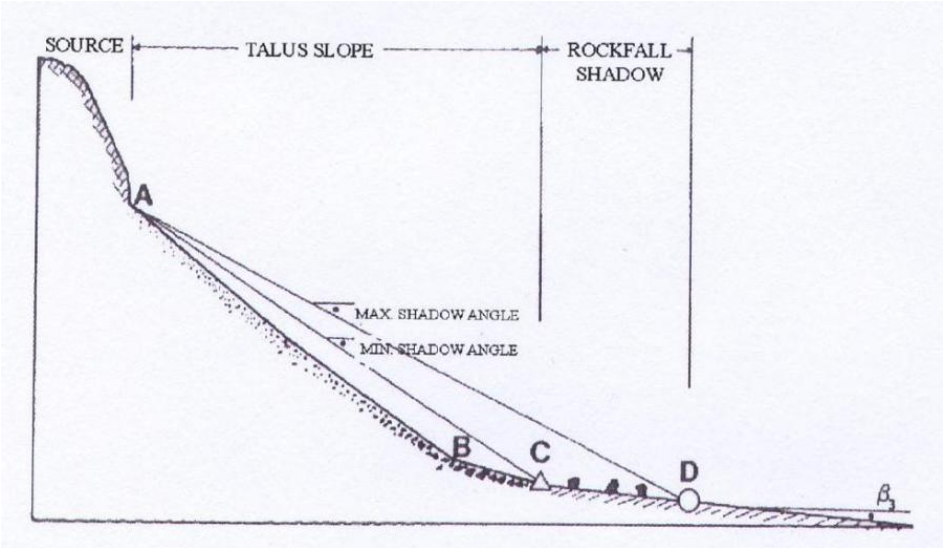


Figure 3-7: Illustration of a characteristic rockfall path profile (Hungr 1989).

4. Standard statistical methods

Two significant aspects of any distribution are its center and the spread of observations. If we have a normal population, the center is often expressed by its mean μ and its spread by the standard deviation σ. Confidence intervals and tests of significance (F-tests) however, are based on the sample mean $\bar{x}$, which is an estimate of the unknown μ (Moore 2003). In the same way, the sample standard deviation functions as an estimate of the population standard deviation σ. This chapter intends to describe some of the central methodology with respect to the utilization of t-procedures for inference about sample means. In addition the methodology concerning handling and influence of deviate observations are presented. Last, some statistical procedures utilized independent of the distribution form (non parametric methods) are presented.

4.1 Sampling and robustness of t-procedures

Suppose we draw a random sample of size n with a mean μ and variance σ2/n in a normal distribution. The variance may then be used to make inferences on the population mean μ. If we do not know the population variance we can compute the estimated variance s^2 of the sample by using the formula

$$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} \qquad (4.1)$$

This statistic (4.1) is also known as the mean square. In the numerator we have the sum of squares (S_{xx}) of the observed values from the point estimate of the mean. In the denominator we have the degrees of freedom which is equal to the number of elements in our sample (n) subtracted by the number of point estimates of parameters used in the sum of squares.

In this case we only have one estimate $(\bar{x})$ which is an estimate of μ. As a result the degrees of freedom are equal to n – 1(Freund 1998, Moore 2003).

The one sample t statistic for a sample is known as:

$$t = \frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n}}} \quad (4.2)$$

This expression represents the student t or t-distribution. In the numerator we have the mean $(\bar{x})$ subtracted by the population mean μ. In the denominator we have the square root of the mean square calculated at (4.1) divided on the sample size. The result of the denominator is referred to as the standard error of the mean $SE_{\bar{x}}$ (Moore 2003). This is again dependent on the degrees of freedom used in computing the mean square. Consequently there is a different t-distribution for each sample size and a particular t-distribution will be influenced by the degrees of freedom (Freund 1998).

In most cases the one-sample t-procedure may be safely used for when n≥ 15 unless an outlier or clearly marked skewness is present. When the samples are large that is n ≥40, the t-procedures may be used even for clearly skewed distribution (Moore 2003) .

The results of the one-sample t-procedures are sensitive to the normality of the sample. This means that it is only exactly correct when the distribution is normal. Since perfectly normal distributions never occur, the robustness of the t-procedure is crucial for the accuracy of our statistics (Moore 2003).

It is crucial to understand the relationship between the size of the sample (n) and the estimated variance. This connection may be observed by plotting random numbers from a normal distribution . If we use a distribution with a mean of 10 and a standard deviation of 2, we may determine the variance of sample sizes between n∈ [3,41]. While the sample size for the population is s^2=4, we find that the independent variance range for each of the samples increases significantly with a smaller sample. The spread is also notable for larger samples as n=40 seem to vary between 2.5 and 6. That being said, it is not as significant as for the declining samples.

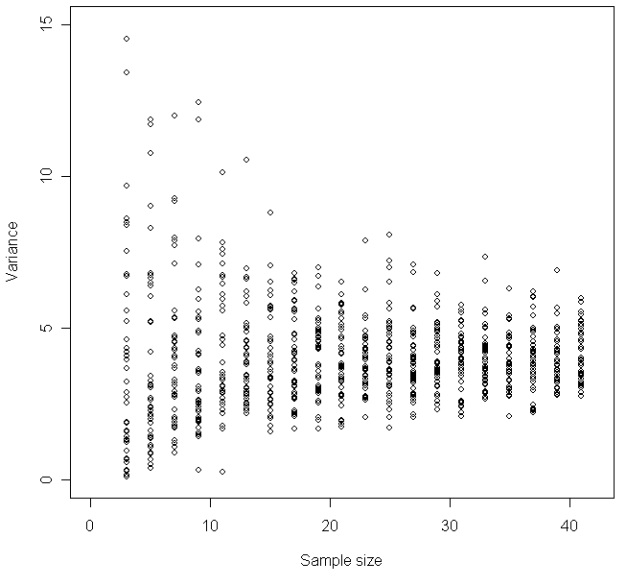


Figure 4-1: Plot displaying the decreasing variance with increased sample size (Crawley 2005).

From this we infer that as the sample size declines the variance will increase significantly. As a consequence we want to have a sample that is as large as possible to avoid variability in variance. If not, it may have serious consequences for our estimation and hypothesis testing. This is demonstrated in figure 4-1 which display a smaller variance with increased sample size (Crawley 2005).

The reason for the sensitivity of the t-procedure is that the mean and standard deviation is not resistance against outliers. That being said, an increase in sample size will improve the accuracy of the P-values and the critical values of the t-distribution. This is because of the central limit theorem which states that the mean from a large sample can be closely approximated to a normal distribution, regardless of the distribution of the population from which the sample is taken from. Because of this we do not need to be as concerned about the normality of individual observations when the sample is large. In addition as the size n of the sample increase, the sample standard deviation (s) move towards the population standard deviation σ. This fact is related to the law of large numbers. Subsequently, the accuracy of the standard deviation (s) in a sample, will be closer to the standard deviation of the population (σ), despite occurrence of non-normality in the distribution (Moore 2003).

The statistical t-test is used to detect deviations from the null hypothesis. For this thesis where we use a regression approach, the null hypothesis is H_0: $\beta_1=0$, whereas the alternative is H_a: $\beta_1\neq0$. We use these two hypothesizes against each other and determine the significance of the regression coefficient based on the probability that the null hypothesis is true.

To calculate the power of the t-test we assume a fixed level of significance, often $\alpha = 0.05$. Thus for very low probabilities under the given α-value we reject H_0, as the alternative value H_a is significant for our analysis (Moore 2003).

The density curves for the t-distribution is similar in shape to the standard normal distribution. This means that they are symmetric about 0 and are bell shaped. That being said, the spread of the t-distributions is a bit greater than that of the standard normal distribution. This is because of the extra variability in the sample caused by replacing the fixed standard deviation of the population with the standard deviation (s) of the sample. Consequently the t-distribution shows more probability in the tail and less in the center than the standard normal distribution.

When the degree of freedom k increase, the t(k) curve will approach the curve representing the standard normal distribution. This is due to the fact that s approaches σ with increasing sample size (Moore 2003).

4.2 Outliers and influential observations

Observations that are noticeably different or atypical from the rest of the observations of a data set are known as an outlier. The occurrence may be with respect to a response variable, the independent variables or both (Freund 1998).

An observation that causes regression estimates to be significantly different from what they could be if the observations were cut out from the dataset is called an influential observation. In order to fully understand the influence of an outlier we need to consider the leverage. More specifically deviation points in independent variables are said to have high leverage and are often referred to as leverage points. Although some observations might look like outliers or have high leverage it doesn't mean that they necessarily are influential. However influential observations are usually outliers and possess a high leverage (Freund 1998).

There are many ways in detecting an outlier. One method for independent variables is to use the 1.5*IQR criterion. In this method we use the interquartile range which is the difference between the first (Q_1) and the third (Q_3) quantile. This is a method were an observation is called a suspected outlier if it falls more than 1.5*IQR above the third quartile (Q_3+IQR) or below the first quartile (Q_1-IQR) (Moore 2003).

How we handle outliers is a matter of judgment. Some observations may point to the atypical nature of some observations (Freund 1998). In this respect it is important to note that it doesn't have to be due to an erroneous data measurement in a research study. Consequently we should be careful about leaving out points, just for the convenience of our analysis.

4.2.1 *Effects of an outlier*

It is important to understand the influence an outlier exerts on the statistical estimates by its position. This may be illustrated by two different outlier scenarios.

1. When x=5,y is increased 10 units
2. When x=10, y is increased 10 units

In both scenarios the observed value of the dependent value is increased by over three standard deviations. The difference lies in the location of the outlier. In the first scenario the outlier is in the middle of our observation, whereas in the second (right) location it is in the high end of the plot (refer to fig.4.2).

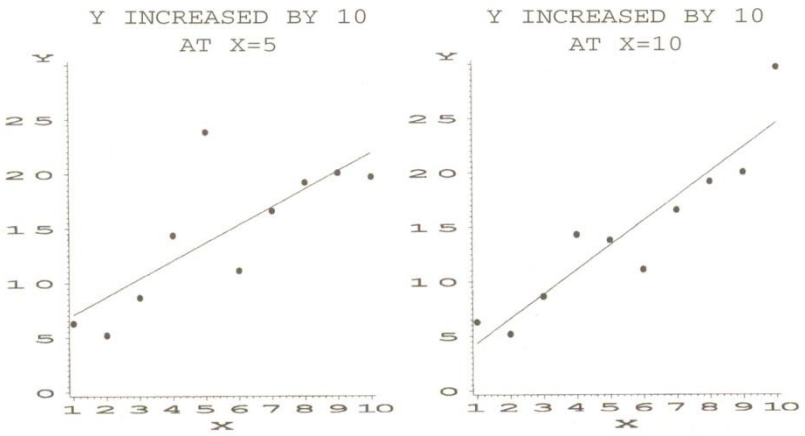


Figure 4-2 : Plot to illustrate how the location of the outlier affects the regression line (Freund 1998).

In order to understand the influence of an outlier, Hooke's law in physics provides a beneficial analogy. This law states as follows " The energy in a coil spring is proportional to the square of the distance the spring is stretched". In effect a hypothetical coil spring stretched from each point to a rigid rod representing the regression line, will produce a least squares line which is achieved by the equilibrium position of the rod (Freund 1998).

As a consequence of this analogy the outlier in the left plot (fig.4-2) exerts a balanced pull on the regression line by lifting it slightly and in effect increasing β_0 while leaving the slope β_1 more or less unchanged. In contrast the outlier in the right plot of fig.4-2 is pulling the line in only one end towards itself, thus it provides more leverage than the outlier in the middle. In effect the estimated slope β_1 is much more influenced by this point. This is also the reason that the means square error in scenario 2 has not increased as much. Thus we see that the location of the outlier has a significant effect of the parameter estimates. For this reason it is not merely the outlier, but also the leverage that defines an influential point (Freund 1998).

4.2.2 *Interpreting residuals and influence of observations*

The significance of an outlier is caused by the combined effect of the magnitude and the leverage of the independent variable. This is referred to as the influence, and is ultimately an estimate of the effect the outlier has on the parameter estimates and hence on the estimated response y (Freund 1998).

In regression analysis, deviations are identified by looking at the scatter of data points about the regression line. This noise or variability about the regression line are called residuals and represent the difference between an observed value and the value predicted by the least squares regression line. Because the residuals display the fit of the regression line a closer examination helps us to assess how well the line describes the data (Moore 2003).

In order to evaluate the fit of a regression line a residual plot which exhibit the regression residuals against the explanatory variable is a valuable tool.

However, this in itself is not enough to determine the influence of a variable. Figure 4-3 shows the corresponding residual plots to the plots in fig.4.2. It reveals that the outlier in the left plot is more noticeable. For the other residual plot belonging to the high leverage observation (x=10), the larger magnitude leads to a deviation from the line which is not much larger than for at least one of the other observations.

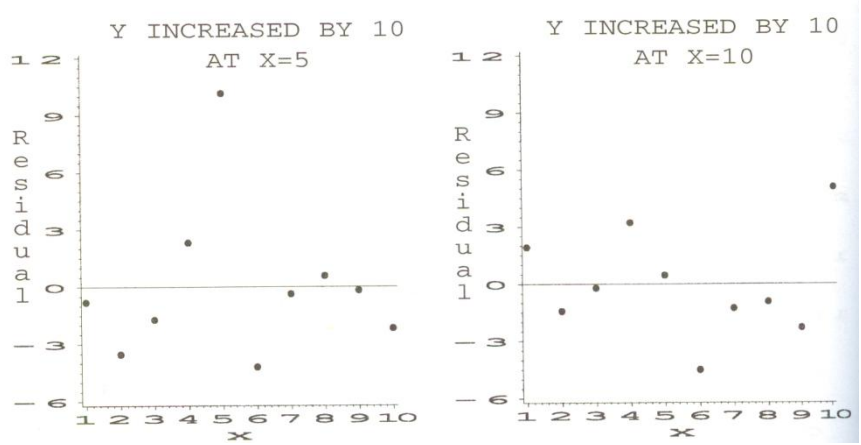


Figure 4-3: Display of the residual plots to the regressions of fig.4-2 with different leverage.

The influence of an observation may be evaluated by leaving out one observation or deleting the residual. In this principle the model utilizing all observations are compared with various model estimates obtained by leaving out each observation separately. The influence of each observation will then be determined according to the magnitude of the difference between the individual observation and the complete model. In this way an observation with a large difference poses a high influence on the estimated model.

A widely used influence statistic is the standardized difference in fit abbreviated DFFITS. This value is calculated in the following way

$$DFFITS = \frac{\hat{\mu}_{y|x} - \hat{\mu}_{y|x,i}}{Standard\ error} \quad (4.3)$$

Where the $\hat{\mu}_{y|x}$ is the estimated mean of the response variable, while $\hat{\mu}_{y|x,i}$ is the estimated mean response variable using the model with observation i left out. The standard error is the standard error of the numerator.

Another type of diagnostic plot that is effective in identifying influential observations is Cook's distance. The formula for this statistic is essentially $(DFFITS)^2/(m+1)$, where m denotes the number of independent variables in the model. The method is not quite as sensitive DFFITS. However as a result of the squared values the potential influential observations are usually more prominent in the plot (Freund 1998).

4.3 Statistical significance of the model

The significance of the explanatory variables will be based on the statistical t-procedures. An equivalent type of this test to judge the overall significance of the model is the ANOVA F-test. As with the t-test the hypotheses are H_0: $\beta_1 = 0$ or H_1: $\beta_1 \neq 0$. When H_0 is true, this statistic has an F distribution with 1 degree of freedom in the numerator (df_M) and n-2 degrees of freedom in the denominator (df_E). When the sum of squares for the model (SSM) and the sum of squares of error (SSE) is divided on the respective degree of freedom we get the difference in variance.

Subsequently the test compares the variances of the model in the following manner

$$F = \frac{MSM}{MSE} = \frac{SSM/df_M}{SSE/df_E}$$

MSM is the variance explained by our model (the mean squares of the regression), while MSE is the variance of the error between the predicted $\hat{y}$ values and the real y-values (mean square of the error).

Thus, the F-statistic divides the variance in the model that is explained, with the variance in the model that is not explained and relates the result to a corresponding probability parameter of the model performance.

This value is the probability that a random variable which has the F (1, n-2) distribution is greater than or equal to the calculated value of the F statistic (Moore 2003).

4.4 Non parametric methods

When we have a clearly non-normal distribution there are some main methods which may be applied. One of this is to transform the skewed data so that the distribution is symmetric and as close to normal as possible. In effect confidence levels, p-values and t-procedures for the transformed data will have a relatively high accuracy-level even for moderate sample sizes (Moore 2003).

Another strategy is to use distribution free inference procedures, referred to as non parametric methods. These methods do not presume any normality or other specific form of the population distribution (Moore 2003). Thus, they contrast to normal parametric tests where it is assumed that the calculated test values have a distribution shape of a known form.
In this respect, the parametric tests all have an underlying assumption of normality. However, we can invoke the central limits theorem to validate utilization of these methods even when the samples are not normal. The conditions is that our sample is large and that the population does not differ to much from normality (Davis 2002).

Non parametric tests use information of a lower rank, rather than metric data required by conventional tests. That is nominal observations where the numbers are simply labels, or ranked observations (ordinal). If the sampled population follows the assumed distribution, a non parameteric test is generally not as precise as the parameteric method. If it deviates from the assumed distribution they are usually more precise (Davis 2002) .

4.4.1 *Bootstrapping*

Bootstrapping is a non parametric computer based procedure for assigning measures of accuracy to statistical estimates (Efron 1993). The basic design is to proceed in the notion that the sample is a valid representation of our population. The bootstrap procedure is then commenced by taking several samples from it. Each of these samples are referred to as a resample (Moore 2003).

The key idea of the inference method is sampling with replacement. As an example, imagine that we have a sample with only five observations which we label as A, B, C, D, and E. We then proceed to draw one observation randomly and record the label. After this is recorded we put the observation back in the sample. This random selection is repeated up to 1000 times or more until we get a final record of the sample. If we want to assess the statistical accuracy of the data basis, we can take N bootstrap resamplings and compute the statistics from each bootstrap resample (fig.4-4). The values of the bootstrap statistics may then be utilized to assess the statistical accuracy of the original sample statistic. It is important to note that since bootstrapping is based on random samples it will not give absolute results, but rather a estimated values (Teknomo 2006). Even so, the fact that we can do statistical estimates of our data without knowing the distribution of the population makes bootstrapping a practical method.

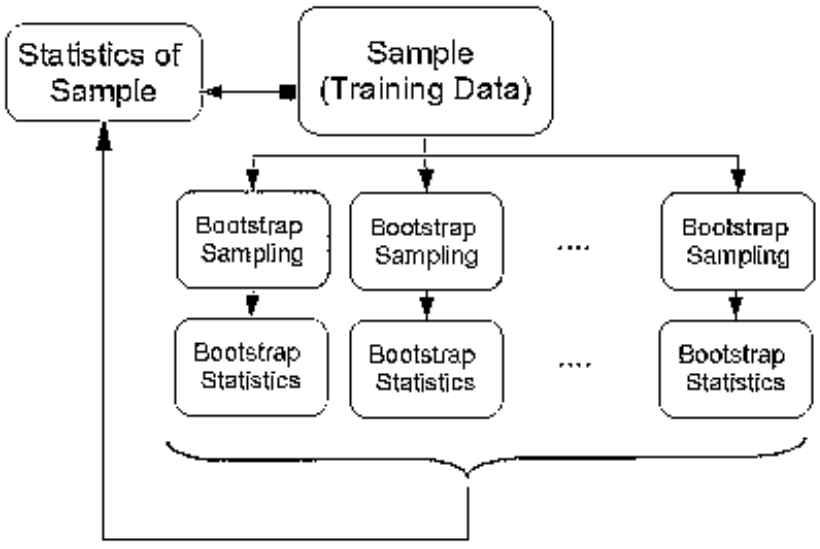


Figure 4-4: Flow chart showing the basic principles of the bootstrap method (Teknomo 2006).

4.4.2 *Kolmogorov-Smirnov test*

Kolmogorov-Smirnov is a non parametric test which has been extensively used in geological context. It is primarily used to test the goodness of fit between models. An advantage with the method is that it compares cumulative distribution functions directly. Thus, it is not necessary to group observations into arbitrary categories (Davis 2002).

The general test is performed by selecting a sample from an unknown population. Then the goodness of fit to a hypothetical model of a specific population is tested. The sample in addition to the hypothetical model is plotted together in a cumulative form and scaled so that their respective cumulative sums are equal to one; in effect expressing both as cumulative probability distributions. Next the greatest difference between the two distributions is found. This maximum difference is referred to as the Kolmogorov- Smirnov statistic (D). Thus we have

$$\boldsymbol{D = max\ |CDF - EDF|}$$

Where CDF denotes the theoretical cumulative distribution function and EDF denotes the empirical cumulative distribution function. D can be used for both one tailed and two tailed null hypothesis. In the case of the two tailed analysis the classes of the distribution from which the sample is derived are equal to the sample in the hypothetical model for all x-values. The one tailed null-hypothesis states that all classes of the sample distribution are equivalent or less than the values of the hypothetical model. In which case the maximum positive difference become the test statistic.

Conversely, we may have that all classes of the sample distribution are equivalent or greater than the values of the hypothetical model. In most instances the Kolmogorov-Smirnov test is utilized when the hypothetical model can be completely specified (Davis 2002).

4.5 Transformations

One of the most common transformation methods is the logarithmic transformation. The logarithm will have a tendency to pull the right tail of a distribution. Thus this is a potential method to correct a right skewness in a distribution.

The main benefit of analyzing the transformed data is that the utilizations of the statistical procedures which are based on normal distributions are better justified and more exact. The disadvantage is that the confidence intervals change with the transformation and may not be recovered to the original distribution (Moore 2003).

5. Regression analysis

One method to examine the problem of rockfall runout range is to use linear regression models. This type of statistical models is frequently accompanied by probabilistic statistics as confidence intervals and cumulative probabilities. The intention of utilizing linear regression models is to observe if there exists any correlation (r) between the different geometrical parameters for rockfall.

Nevertheless, linear regression as a statistical method for inference has some limitations that should be taken into account when evaluating the validity of the results. First and foremost, it is important to note that an indication of a regressive relationship does not automatically imply that *x* causes *y*. In other words, a statistical regression model is never exact; rather it provides us with estimated values which fit the data basis more or less.

The intention of regression analysis is not to make inferences on differences among the means of populations, but to make inferences about the relationship of the mean of the response variable to the independent variables. Subsequently, a relation between two variables does not by itself determine a result. It must also be demonstrated that no other factor could give the result (Freund 1998).

In addition, it is not recommended to use an estimated regressive relationship for extrapolation. Even though the model may fit the sample well, such statistical inference is erroneous since there is no evidence that the model is appropriate outside the range of the existing data (Freund 1998).

5.1 Simple linear regression analysis

The statistical method of simple linear regression is used to describe how a dependent response variable *y* changes with an explanatory independent variable *x*. A regression line determines the relationship between the two variables by using the principle of the least-squares method, while the deviation of points outside the line describes the residuals. The regression line is the line that makes the sum of the squares of the vertical distances as small as possible. The square of the correlation (r^2) is the fraction of variation in the y-values that is explained by the least-squares regression of *y* on *x* (Moore 2003). The equation for the least squares regression for a population (5.1) is written as follows (Freund 1998)

$$\mu_y = \beta_0 + \beta_1 X + \in \quad (5.1)$$

Slope:

$\beta_1 = r\frac{Sy}{Sx}$ (5.2)

Intercept:

$\beta_0 = \bar{y} - \beta_1\bar{x}$ (5.3)

Where β_1 , (defined in equation (2)), denotes the slope of the explanatory x-variable, whereas β_0 ,(defined in equation (3)), denotes the intercept on the y-axis. These parameters are referred to as regression coefficients and compose the deterministic straight line relationship in the model.

Statistical models for simple linear regression assume that for each value of x the observed values of the response variable y are normally distributed about a mean that is dependent on x. In this respect the μ_y represent these means. Our intent is to establish how many means (μ_y) that changes with respect to x. When plotted against x in simple linear regression it is assumed that they all lie on a line according to equation (5.1). Furthermore, the models assume that the variation measured by the standard deviation σ, is the same for all values of x.

The slope parameter (β_1) is composed of the correlation coefficient (r) multiplied with the standard deviation (s) for *x* and *y* respectively. The correlation parameter (5.4) estimates direction and strength of the linear relationship between two quantitative variables. Correlation between *x* and *y* is equal to:

$r = \frac{1}{n-1}\sum\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_{i-\bar{y}}}{s_y}\right)$ (5.4)

As a result a positive r will indicate a positive trend between the variables, while a negative r indicates a negative trend. A correlation value near 0 indicates a very weak linear relationship, while a value close to -1 or 1 indicates a strong correlation where the points lie close to a straight regression line (Moore 2003).

The square of the correlation (5.5) indicates the fraction of the variation in y that is explained by the least squares regression of *y* on *x*. Thus r^2 is a measure of how successfully the regression explains the response. If there are no deviations between the predicted and observed response variables the correlation is 1 or -1 (Moore 2003).

$r^2 = \frac{variance\ of\ predicted\ values\hat{y}}{variance\ of\ observed\ values\ y}$ (5.5)

The residual part of the model is indicated by the $\in$. This term explains the variability of the responses around the mean, which is the variance (σ^2).

A response y is then the sum of its mean and the chance deviation ∈ from its mean. The residuals may then be defined as (5.6, 5.7) (Freund 1998):

∈ = observed response variable – predicted response variable

$$= y_i - \hat{y}_i \qquad (5.6)$$
$$= y_i - \beta_0 - \beta_1 x_i \qquad (5.7)$$

Where $\in_i$ comes from a population with mean 0. The scatter between two variables signifies the error of the linear regression line. The method of least squares chooses the line that makes the sum of the squares as small as possible. In order to find this line we must find the intercept β_0 and slope β_1 that minimizes the deviation for the given variables. We will then have the following minimizing equation (Freund 1998):

$$\sum\left(y - \hat{\mu}_{y|x}\right)^2 = \sum(y_i - \hat{\beta}_0 - \hat{\beta}_{1xi}\ )^2 \quad (5.8)$$

Where the notation $\hat{\mu}_{y|x}$ signifies that the estimated mean of the variable y is dependent on the value of x. The coefficient value that minimizes the sum of squared deviations for a given set of sample data may be estimated by applying the normal equations:

I. $\hat{\beta}_0 n + \hat{\beta}_1 \sum x = \sum y \qquad (5.9)$

II. $\hat{\beta}_0 \sum x + \hat{\beta}_1 \sum x^2 = \sum xy \qquad (5.10)$

By solving these two equations with two unknowns we can derive a formula for the estimated regression coefficients. This will result in the following relations:

$$\hat{\beta}_1 = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - (\sum x)^2/n} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{Sxy}{Sxx} \quad (5.11)$$

Which is the same as the corrected sum of products of xy (s_{xy}) divided on the corrected sum of squares for the independent variable x (s_{xx}) (Freund 1998). This may be found by using a calculator or software.

5.2 Quantiles and probability

We can describe the spread or variability of a distribution by using several percentiles referred to as quantiles. The quantile of a distribution is a number x_p such that a proportion p of the sample values are less than or equal to x_p. Thus, the 25th quantile, also referred to as the 25th percentile is a variable x_p such that 25% of the values of the variable (x_p) fall below that variable (Moore 2003).

The cumulative distribution function (cdf) is the probability that the variable has a value which is less than or equal to x. Thus, we have the following relation:

$$F(X) = P[X \leq x] \qquad (5.12)$$

Where $F(X)$ denotes the cumulative distribution function (cdf). $P[X \leq x]$ denotes the probability that the observation has a value which is less than, or equal to an observation x in the data basis.

Conversely the survivor function S(X), which follows what is referred to as the complement rule, will then be the probability that the variable has a value which is larger than x. Thus the exceedance probability for an event x follows the rule (Moore 2003):

$$S(x) = P[X > x] = 1 - F(x) \quad (5.13)$$

We see that the results of these simple probability rules imply that the probability that an event occurs is always equal to 1 minus the probability that the event does not occur. In this respect the complement rule is very useful since in many problems it is simpler to calculate the probability that an event does not occur, than to calculate the probability that the event occurs.

5.3 Confidence and prediction intervals

The confidence intervals may be used for specific statistics like means or regression lines. It establishes a range of values around the desired statistic where the "true" statistic can be expected to be located within a given certainty. The confidence interval is regularly written in the form:

estimate ± margin of error

In the case of the mean of a standard random sample, $(\bar{x})$ is the estimated guess for the value of our unknown population parameter. The margin of error shows how accurate we believe our guess is, based on the variability of the estimate. The confidence level tells us how confident we are that the procedure will derive the true population mean μ.

Overall there are two principal things to know about a confidence interval:

1. It is an interval (a, b) where a and b are numbers computed from the data.
2. It has an attribute called a confidence level which provides the probability that the interval covers the parameter.

In a normal distribution what we call the 68-95-99.7 rule states that the probability is about 0.95 or 95% that $(\bar{x})$ will be within two standard deviations of the population mean μ
In a linear regression perspective our intent will often be to estimate the confidence interval for the mean response of the population (Moore 2003).

In this respect, a specific value of an explanation variable x, (e.g. x^*), the mean of the response y in the subpopulation will be given by:

$$\mu_y = \beta_0 + \beta_1 x^* \quad (5.14)$$

In order to estimate the mean from the sample basis, we substitute the estimates b_0 and b_1 for β_0 and β_1. In effect this yields:

$$\hat{\mu}_y = b_0 + b_1 x^* \quad (5.15)$$

The subsequent confidence interval for μ_y estimates a margin of error based on the standard error$\left(SE_{\hat{\mu}}\right)$. In this respect the standard error of $\hat{\mu}$ is equal to:

$$SE_{\hat{\mu}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \quad (5.16)$$

Where s is the estimated standard deviation of the sample and $\bar{x}$ the estimated mean of the sample. By utilizing $SE_{\hat{\mu}}$ as a margin of error together with a distributional t-statistic, a computer program can derive the confidence interval for the mean response corresponding to each of the x-values in the sample basis, or any value x^* of the explanatory variable. With these conditions, a set confidence level (CI) for the mean response μ_y when x takes the value x^* will be equal to:

$$\hat{\mu}_y \pm t^* SE_{\hat{\mu}} \quad (5.17)$$

Where t^* is a value from the t(n-2) density curve with an area between $-t^*$ and t^*.
In a graphic display the confidence lines will be narrowest for x^* values around the distribution mean of the observed values and widen as x^* moves away from $\bar{x}$. However, as stated earlier we should avoid extrapolation for extreme values which are far away from our observation basis.

Prediction intervals are in many ways similar to regression confidence intervals. However, as it is harder to predict one individual value as opposed to the mean, the margin of error is larger. The predicted future response $\hat{y}$ for an individual case with a specific value x^* of the explanatory variable is equal to:

$$\hat{y} = b_0 + b_1 x^* \quad (5.18)$$

Which is the same estimate as the expression for $\hat{\mu}_y$. Thus, the fitted least squares line is used both to estimate the mean response for any explanation variable x^* in relation with the mean response, as well as for prediction of a single future response. Furthermore, in the same way as for the confidence intervals, the prediction intervals include a margin of error as an indication of its accuracy. The form of the prediction interval is comparable to that of the confidence interval. Even so, it has the important difference that it uses the standard error of the $SE_{\hat{y}}$, which includes both the variability caused by the fact that the least-squares line is not exactly true to the regression line, in addition to the variability of the future response variable around the sample mean. The following formula is used to derive the standard error for prediction of an individual response (Moore 2003):

$$SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \qquad (5.19)$$

Following the general outline of confidence intervals, the prediction interval for a future response variable y from the subpopulation corresponding to x^* becomes:

$$\hat{y} \pm t^* SE_{\hat{y}} \qquad (5.20)$$

Where t is a value from the t(n-2) density curve with an area between $-t^*$ and t^*. In a graphical display the upper and lower line of the prediction intervals will be further from the least squares line than the line of the confidence limits. Thus, the interval for a single future observation is larger than an interval for the mean of its subpopulation (Moore 2003).

5.4 Multiple regression

In multiple regressions we use more than one explanatory variable to predict a single response variable. However, inference methods used in simple linear regression such as correlation and least squares regression are also applied here (Moore 2003).

The general form of the multiple linear regression is written as an extension of the simple linear model in the following form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \qquad (5.21)$$

Where y denotes the dependent variable, β_0 denotes the intercept and $\beta_n x_n$ signifies the corresponding n regression coefficients (Freund 1998).

The inclusion of several explanatory variables creates an interrelation between the variables. Thus the actual effect of each of the explanatory variables is dependent on what is occurring with the other explanatory variables. In this way the overall effect attached with each of the explanatory variables is an estimate of the average change in the response variable.
The changes will then be associated to that particular explanatory variable, while the other explanatory variables remain fixed. This is the common interpretation for a regression coefficient in a multiple regression model (Freund 1998).

There are several important issues involved in multiple regression analysis which need to be evaluated right from the onset. These are crucial for the validation of the analysis and will be presented in further detail in the following subchapters.

5.4.1 *Risk of over parameterization*

Over parameterization occurs when number of intended estimated parameters surpasses the size of our sample. If we have six explanatory variables to evaluate, we have the possibility of performing 15 two way interactions. Furthermore there are 15 four way interactions, 6 five way interactions and one 6 way interaction. In addition we may have quadratic terms for each of the six explanatory variables.

As a consequence such a case would mean that we have the possibility of estimating about 70 parameters. If we then had a sample of 40, this would mean that we are planning to estimate almost twice as many parameters as there are datapoints, which is not possible. A general rule is that the number of estimated parameters should not exceed n/3 parameters during multiple regression (Crawley 2005).

5.4.2 *Manipulation of explanatory variables*

Before executing multiple regression analyses it is crucial to consider all possibilities among our chosen explanatory variables. If not we may overlook a term that is highly significant for the overall outcome, or response variable.

The most frequent methods to manipulate the explanatory variables in multiple regressions are to use interaction terms and quadratic terms. The latter is done to check for quadratic significance by squaring the explanatory variables.

Interaction on the other hand refers to how the effect on the response of one explanatory variable depends on the level of one or more other explanatory variables. In this way interaction occurs when the effect of one explanatory variable depends on the particular level or value of another explanatory variable. Conversely if the effect of one explanatory variable is constant or remains the same across all levels or values of the other, there is no interaction between the variables (Fitzmaurice 2000).

In this fashion the interaction process checks each and every level and multiplies it with the other independent variable. Thus the manipulation is applied to see if there exists an interaction effect dependent on the level of the two compared explanatory variables.

5.4.3 *Ranking significant variables*

In order to obtain useful results in multiple regression analysis, the variables should be sorted out according to their relevance as predictors for the response variable$y$.
Furthermore, the implication of each variable for our response needs to be established. This objective is achieved by sorting out the significant terms. This may be achieved by application of the probability with respect to each variable's t-value. In this respect we aim for a low probability, which means that we reject the null hypothesis that states that the variable has no implication in the response (Krumbein 1965).

In a stepwise regression procedure the most important single explanatory variable is first obtained. This is then held constant in further regression analysis in order to identify the second most important variable (Krumbein 1965).

5.4.4 *Regression tree*

One efficient way of determining the importance of single variables is to apply the variables to a regression tree. In addition to determining the significance of each variable we acquire the threshold values and see which subordinate variables are most important for high or low means in the hierarchy.

The regression tree uses a weighted method to see whether there exists complex interactions between the explanatory variables (Crawley 2005). More specifically it uses binary recursive partitioning, by utilizing the response variable in the specified formula and then choosing splits from the right side. The numeric variables will then be divided into X < a and X > a. The split which maximizes the reduction in impurity is then selected, the databasis split and the process is repeated. The splitting will go on until the terminal nodes are too small or too few to be split (Ripley 2007).

This process leads to the final outcome of a branched regression tree that displays the threshold values for low and high mean values. In this way the tree-based models provides a valuable tool as they display the most important explanatory variables. Furthermore it gives a perspective in the complexity and interaction of the variables (Crawley 2005).

5.4.5 *Multicollinearity*

The existence of high correlations among the chosen explanatory variables in a regression model is referred to as multicollinearity. This will often occur in the preliminary stages of a multiple regression analysis when a large number of variables are assessed.
Multicollinearity in itself does not affect the overall fit of the model. Consequently it does not influence the model's ability to estimate the response variable or residual variation.
However, it does reduce the effectiveness of a regression analysis, particularly with respect to determining the effect that the various explanatory factor variables exhibit.

In addition the large standard errors of the coefficients contribute to increase the standard errors of the estimated conditional means and the predicted values (Freund 1998).

Another aspect of multicollinearity is that the model containing all variables may not fit the data better than the models with fewer variables. This is because correlated variables may be considered as largely measuring the same phenomenon. Therefore one variable may provide almost the same significance as a combination of several variables (Freund 1998).

6. Evaluation of dynamic models

For the physical modeling of runout range, two different simplified dynamic models are tested and compared. These are a model based on an equation outlined by Kirkby and Statham, formerly adapted and utilized by Keylock (Keylock 1999), and a simple energy equation formerly used and described by A.E Scheidegger. For the latter equation 75 per cent of the velocity was subtracted after impact with the bottom of the cliff. This was done in order to simulate the lost energy during impact with the substrate. Subsequently this model will be referred to as the energy loss model.

A primary difference between the two methods, apart from the use of an energy loss constant, is that the Statham model uses a mean dynamic friction angle as a bulk measure of impact characteristics and movement modes with the surface. The energy loss model on the other hand, employs a separate friction coefficient for each observation. This is derived using the energy line principle on the shadow angle (α). Even so, in order to compare the performance of the two models, a mean friction coefficient has also been applied to the simple energy loss model. Furthermore, the models are used in a lumped mass scenario which implies the following conditions:

- There is no fragmentation of the mass during impact with the substrate.
- The shape of the mass and the angular velocity is not accounted for.
- The average movement of the rock can be treated as that of a sliding mass over a rough surface.

It is important to note that the use of lumped models is generally not as accurate as that of rigorous models. However, the utilization of lumped models is much simpler as they are based on fewer parameters and calculations (refer to section 3.4).

6.1 The Statham model

The approach for the Statham model is based upon the summarized method of Staham and Kirkby, described in section 4.1.2. The model has formerly been utilized on an assortment of the same databasis by Keylock.

For the utilization of the Staham equation the rock follows a lump model perspective with no horizontal velocity during free fall. Furthermore, the utilization implies insignificant energy loss during fall and impact with the bottom of the cliff (Keylock 1999).

The velocity of the boulder is divided into four parts. These are dependent on where on the mountainside the rock is evaluated.

The utilized components of the velocity are the following:

$$v_{b\theta} = \sqrt{(2gH3)} \quad (1)$$

$$v_{t\delta} = \sqrt{(2gH3)}\sin\delta \quad (2)$$

$$v_{b\delta} = \sqrt{v_{t\delta}^2 + 2a_\delta S_2} \quad (3)$$

$$v_{t\gamma} = v_{b\delta}\cos(\delta - \gamma) \quad (4)$$

$$a_\delta = g\left(\sin\delta - \cos\delta\tan\phi'_{\mu d}\right) \quad (5)$$

$$a_\gamma = g\left(\sin\gamma - \cos\gamma\tan\phi'_{\mu d}\right) \quad (6)$$

$$S = -\frac{v_{t\delta}^2}{2a_\delta} \quad (v_{b\delta} < 0) \quad (7)$$

$$S = -\frac{v_{t\gamma}^2}{2a_\gamma} + S_2 \ (v_{b\delta} > 0\ ) \quad (8)$$

Where the subscripts of the velocities (v) and acceleration (a) indicate the assessed part of the mountainside, and the top (t) or bottom (b) of the respective part.

Equation one denotes the free fall velocity above the talus substrate. This is derived using the Newtonian principle of energy conservation where the potential is equal to the height of the cliff (H3). The second equation is an estimation of the velocity at the top of the talus (S2). Moving on, equation (3) denotes the boulder's velocity at the bottom of the talus, while (4) describes the velocity at the top of the terrain beyond the talus. The subscript of the acceleration denotes the assessed slope segment. By reference to equation (5) and (6), this is negative if the force of resistance for the given slope ($g\cos\alpha\tan\phi'_{\mu d}$), exceeds the driving force ($g\sin\alpha$) of the rock. Equation (7) and (8) estimates the runout distance of the rock. The utilization of (7) or (8) is dependent on the velocity at the bottom of the talus($v_{b\delta}$).

The most sensitive parameter when using simplified dynamic models is the dynamic friction angle ($\phi'_{\mu d}$). According to the hazard protection program Rockfor.net uses an empirical dynamic friction angle of 31^{o}, corresponding to a coefficient of 0.6. This empirical value is based on the energy line principle of Heim from estimates conducted at a slope of 38 degrees (Berger 2007).

Equation (3.6) in section 3.1 is a rearrangement of the distance equation (3.5), thus it is based on the modeled mean distance of runout and may not be used on the observed extreme runouts in our sample basis. Alternatively experiments by Statham discovered a linear relation between moving stone sizes (d^{*}) and board stone sizes (d), however to utilize this we need values for the stone sizes in the scree. In addition the relation is limited to angular stones for which the ratio of minor to intermediate axes is in the range 0.6 to 0.8.

As an alternative the mean dynamic friction coefficient of 0.68 derived by Keylock for the sample basis was utilized for as the retardation factor (Keylock 1999). The applied calculation method used to derive this value is not described, however since the sample is based on extreme values, it is logical that the mean value is skewed towards the lower limit of values (Keylock 1999). Furthermore, the fact that the samples are all extreme values suggests that the variance should be low. Thus, the utilization of an average dynamic friction coefficient can be justified.

Although the sample consists of events going beyond the talus slope, the model predicts shorter distances for four observations. As a consequence, these four observations consisting of observation nr. 12, 32, 42 and 44 with S1 values of 32, 26, 20 and 29 meters respectively were dropped from the regression analysis. From the plot in figure 6-1 and the table in figure 6-3, we see that the model tends to underpredict around the 25^{th} percentile. In this respect, observation nr. 109 has a high influence on the percentile. For the original distribution this value is above 100 meters, while for the model it is predicted to a length of 47 meters. Furthermore, the model's general tendency to underpredict in the lower quantile levels is reflected in the model plot (fig.6-1). Also notable is the overall smaller intervals for percentile levels up to about 75. Thus, the spread at different percentiles is less than for the actual sample.

The model yields somewhat better runout estimates than the empirically derived models for values of lower exceedance probabilities. However, from figure 6-4 we see that the overall variance is somewhat lower than in the original sample. Part of the explanation can be found in the neglect of some of the comparatively shorter runout observations that were predicted to not exceed the end of the talus.

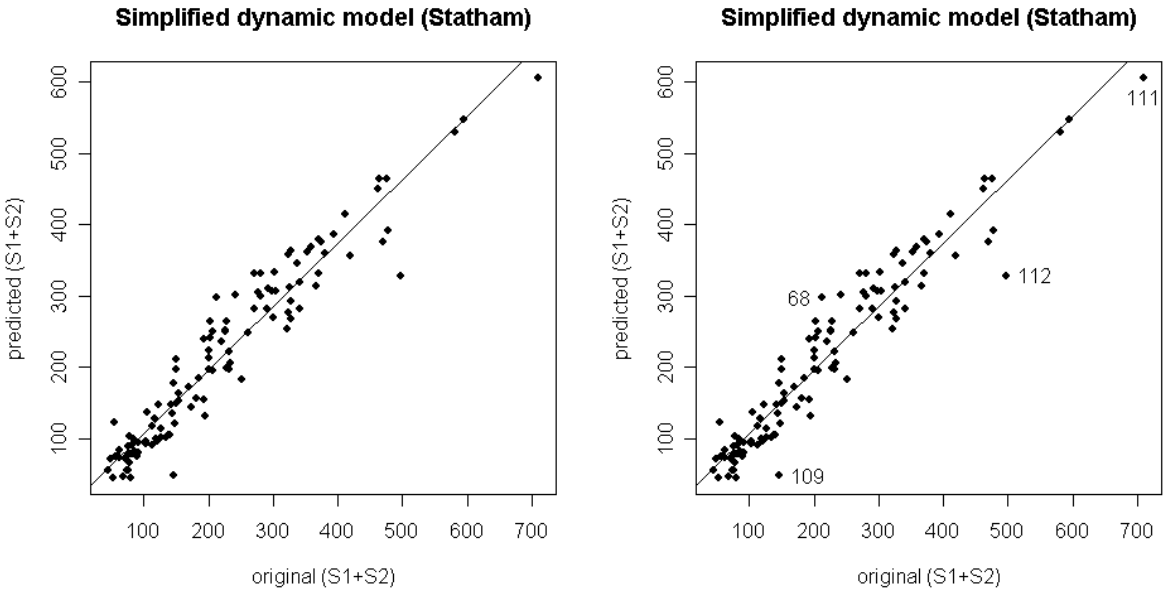


Figure 6-1: Plot displaying the relation predicted (S1+S2) against the original sample basis of (S1+S2).

(S1+S2)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(71.98, 73.24)	(101.5,104.0)	(192.2, 193.0)	(301.2, 303.6)	(469.5, 475.5)
CI=0.25	(71.78, 74.50)	(94.7, 104.6)	(188.1, 201.5)	(299.6, 306.0)	(465.3, 475.5)
CI=0.5	(69.98, 79.01)	(89.7, 114.0)	(184.2, 205.4)	(281.1, 313.2)	(462.5, 478.3)
CI=0.75	(68.69, 80.08)	(86.5, 117.7)	(182.7, 216.5)	(277.4, 323.9)	(460.6, 521.3)
CI=0.95	(65.62, 84.98)	(78.2, 124.7)	(164.7, 236.3)	(275.9, 344.6)	(363.1, 545.6)

Figure 6-2: Table of bootstrapped quantiles (p) and confidence levels (CI) for the original (S1+S2) values.

(S1+S2P)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(74.67, 75.20)	(98.9, 100.5)	(197.2, 197.4)	(305.0, 307.4)	(419.3, 431.2)
CI=0.25	(74.48, 75.85)	(98.5, 100.6)	(197.0, 204.1)	(304.1, 310.6)	(389.4, 444.2)
CI=0.5	(72.80, 77.73)	(97.1, 103.9)	(189.3, 214.3)	(300.4, 314.0)	(387.2, 451.5)
CI=0.75	(69.99, 79.46)	(86.6, 107.9)	(177.1, 226.5)	(294.2, 329.6)	(375.2, 459.0)
CI=0.95	(68.44, 86.92)	(71.2, 111.7)	(154.0, 241.0)	(280.5, 342.2)	(309.8, 469.5)

Figure 6-3: Table of bootstrapped quantiles (p) and confidence levels for (S1+S2) based on Statham model.

Desc. Stats.	st. dev.(s)	variance	mean	median
Sample	134.59	18115.06	215.12	192.8
Model	124.84	15584.00	213.04	197.3

Figure 6-4: Table of descriptive statistics for the dynamic Statham model and the sample basis.

6.2 The energy loss model

For the simplified energy loss model the motion of the rock is approximated to that of a dry mass. Following the energy line principle introduced by Albert Heim for larger dry flow masses, the dynamic friction coefficient is estimated as that of the tangent of an angle. In this case the angle basis is that of the shadow angle (α), as opposed to the more traditional all-embracing β angle (fahrböschung), where the tangent of β is used as a rough approximate to the apparent friction coefficient. In order to test the theoretical value of energy loss formerly proposed by Broili and Hungr (Dorren 2003), 75 per cent of the acquired free fall velocity was subtracted after impact with the talus.

An overall weakness of the model is the fact that the model is highly sensitive to both the dynamic friction coefficient and to the angle of the terrain beyond the talus. As a consequence the runouts which are influenced by a large γ parameter will be prone to an overprediction. Thus, the model is not liable to be an accurate predictor for the observations which have low exceedance probabilities in the independent runout distribution. The model is based on three simple formulas which are dependent on the location of the mountainside.

These are the following:

$$v_{t\delta} = \sqrt{(2gH3)} \sin\delta * 0{,}25 \quad (1) \qquad v_{b\delta} = \sqrt{v_{t\delta}^2 + 2g((H2) - x\mu)} \quad (2)$$

$$S = \frac{v_{b\delta}^2}{2g(\mu - \tan\gamma)} \quad (3)$$

The first formula (1) models the boulder's velocity at the top of the talus slope. In order to simulate the energy loss, 75 percent of the velocity is assumed lost after impact with the bottom of the cliff (H3). This value of energy dissipation is currently used in the program Rockfor.net for calculation of rockfall hazards (Berger 2007), and has also been utilized in GIS models for rockfall prediction (Dorren and Seijmonsbergen 2003).

Equation (2) is the calculation for the velocity at the bottom of the talus segment, where g is the gravitational acceleration constant, H2 is the height of the talus slope, x is the horizontal distance of the talus part (equal to L2) and μ is the dynamic friction coefficient. The selected dynamic friction angle was based on the tangent of the shadow angle (α), that is the straight line between the endpoint of the rock and the beginning of the talus (refer to fig.2-1 for illustration).

As a first estimation a separate value for retardation was employed for every observation. This produced a mean friction coefficient of 0.47 and a relatively small standard deviation of 4.86°. Former literature values give upper and lower limits for μ ($= \tan\phi'_{\mu d}$) between 0.4 and 1.5 (Keylock 1999). Thus, the value is definitely towards the lower end of the spectrum.

That being said, as discussed with respect to the Statham model, our sample basis is considering the extreme runout distances beyond the talus. Subsequently, the utilization of a low dynamic friction coefficient is reasonable.

The model tends to overpredict somewhat for the runouts of low exceedance probabilities (refer to right plot of fig.6-5). In this respect the three longest runouts with ranges above 200 meters (obs. 120,121 and 122) where clearly the most deviate. This is assumed to be a result of the large γ angle which effectively produces a smaller number in the denominator and thus increases the runout distances. However, the overall performance of the model predictions was relatively good. This was proven by a squared correlation value of 0.986 and a relatively low residual model error of 16.27 meters. The results indicate that the conditions in the model with the use of an energy loss factor and the tangent of the shadow angle yields a good first estimate of runout distance. If we merely consider the restitution aspects of this, the result is unexpected and should be investigated further. In effect this suggests that a tangential restitution of around 70-75 per cent during impact with the substrate is a reasonable estimate and that the variance is fairly limited. With reference to equation I in section 3.2 this would mean that the boulder has a negligible rotation in the air before colliding with the bottom of the cliff. The other important suggestion of the result is the significance of the shadow angle as a basis for coefficient of restitution. The accuracy of the model is in many ways a confirmation that the length and slope of the talus plays an important factor in the retardation of the boulder.

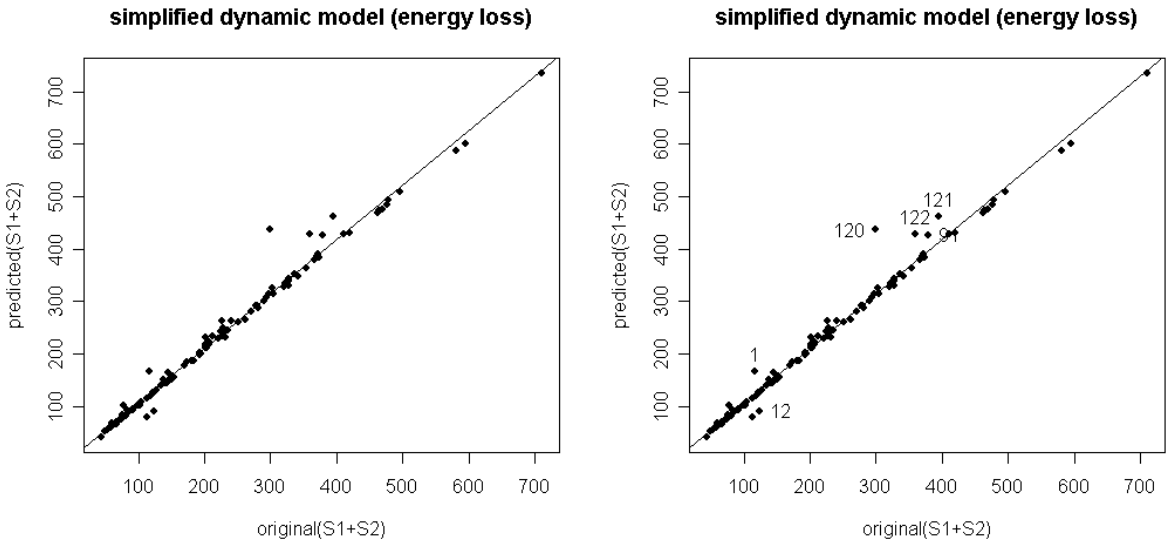


Figure 6-5: Plot displaying the simplified energy loss model.

For the sake of comparison with the Statham model, the energy loss model was used with the calculated average friction coefficient (μ= 0.47). This produced two runout distances that stopped before the end of the talus. These were observation number 32 and 42 with original runout distances of 26 and 20 meters respectively. As expected, the model produced a larger variance than for the Statham model (fig.6-6).

This is because the model is much more sensitive to the dynamic friction angle than in the Statham model, where the coefficent is a subcomponent in the determination of the acceleration at different parts.

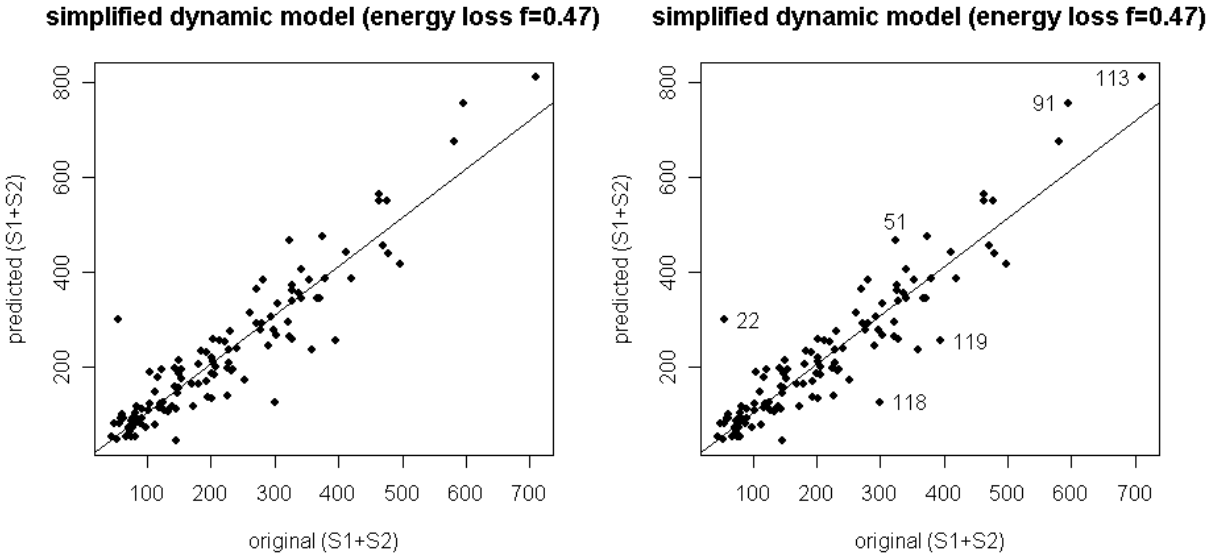


Figure 6-6: Plot of the energy loss model with an average friction coefficient of 0.47.

When comparing the bootstrapped values of the sample (fig.6-7) with the model results utilizing an average friction coefficient (fig.6-8) we see that the model is significantly deviate at the different percentile levels. This is further reflected in the plot which displays a significant spread at both sides of the regression line. Subsequently the model does not show any overall trend of under prediction or over prediction.

(S1+S2)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(71.98, 73.24)	(101.5,104.0)	(192.2, 193.0)	(301.2, 303.6)	(469.5, 475.5)
CI=0.25	(71.78, 74.50)	(94.7, 104.6)	(188.1, 201.5)	(299.6, 306.0)	(465.3, 475.5)
CI=0.5	(69.98, 79.01)	(89.7, 114.0)	(184.2, 205.4)	(281.1, 313.2)	(462.5, 478.3)
CI=0.75	(68.69, 80.08)	(86.5, 117.7)	(182.7, 216.5)	(277.4, 323.9)	(460.6, 521.3)
CI=0.95	(65.62, 84.98)	(78.2, 124.7)	(164.7, 236.3)	(275.9, 344.6)	(363.1, 545.6)

Figure 6-7: Table of bootstrapped values. Performed on original runout sample with 1000 replicates.

(S1+S2P)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(79.48, 79.77)	(111.4, 112.3)	(192.9, 193.5)	(292.7, 297.6)	(479.5, 483.2)
CI=0.25	(78.28, 80.76)	(111.1, 114.0)	(192.5, 196.4)	(287.5, 299.9)	(408.8, 492.8)
CI=0.5	(77.88, 82.76)	(110.4, 116.6)	(188.7, 201.6)	(273.0, 312.7)	(408.5, 502.5)
CI=0.75	(75.38, 86.27)	(106.7, 121.8)	(181.5, 212.0)	(251.4, 321.4)	(407.8, 516.7)
CI=0.95	(68.16, 94.58)	(99.4, 132.3)	(162.7, 221.9)	(243.9, 335.4)	(287.0, 569.3)

Figure 6-8: Table of bootstrapped values of energy loss model (fric.coef=0.47). Performed on model values with 1000 replicates.

7. Statistical inferences

A basic methodology in all statistical procedures is to use facts about a sample to estimate the truth about the whole population. If the sample is relatively large this may expand the current conclusions about the phenomenon of rockfall in general. To infer this knowledge we use our sample basis of 122 independent rockfalls collected from various locations in Norway by Ulrik Domaas. One of the central rules in all statistical deductions is to begin the analysis by simply observing or exploring the distributional aspects of the variable we wish to gain an understanding of.

By using this methodology we will have a solid knowledge of what the original data tells us, and an instinct towards which variables or individual values which have a relatively greater contribution to our new statistical derivations.

The linear regression results are produced by analyzing the correlation between different variables that are believed to be related to each other. The produced results for each analysis include the t-statistics, and corresponding probability, together with the F-statistic of the model performance. It is important to note that all predicted runout values (S1), except for the case of the multiple regression analysis, are added to the length of the talus hill (S2). As a direct consequence the actual correlation value will be increased for each model analysis.

7.1 Exploratory data analysis

Exploratory data analysis (EDA) is a statistical approach of data analysis that employs mostly graphical procedures to infer knowledge on the population a sample is taken from (Krumbein 1965). The intention behind the analysis method is to suggest the kind of probability density function (pdf) which fit the sample we want to model.

The most crucial factor when determining the risk at a site is to establish the potential rockfall range. In this respect it is reasonable to see how the distribution of our sample is with respect to the quantitative variable of runout (S1) beyond the talus slope.

One practical method of examining this is to apply the boxplot function (fig.7-1). This will create a box around the median and borders where the 25th and 75th percentile is located. The whiskers will extend to the smallest and largest values that are not suspected outliers, while observations that are more than 1.5*IQR outside the box are plotted as individual points.
In addition to this, the histogram is an important tool for all distributions. The reason for this is that it depicts the frequency of the observations, which ultimately leads to a better overview and understanding of our dataset. When displayed graphically as a boxplot and histogram respectively, the distribution reveals four leverage points or outliers. All of these are above a runout range of 200 meters.

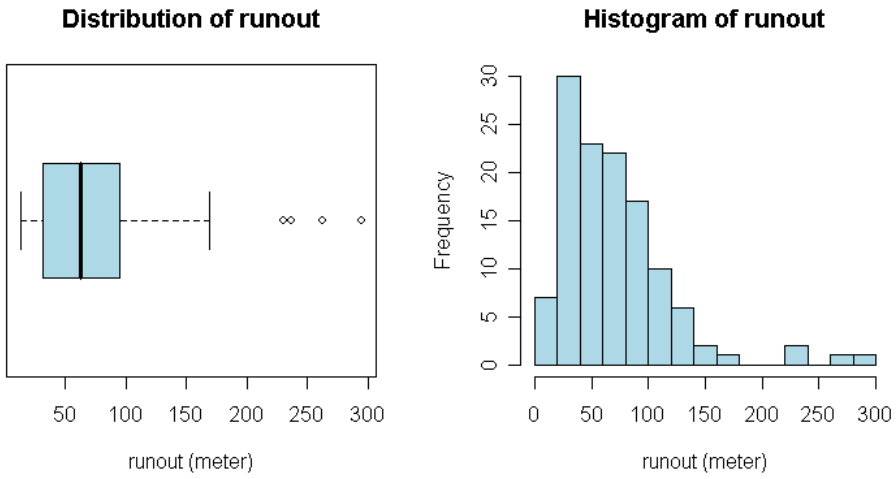


Figure 7-1: Figure displaying the distribution of independent variables for runout (S1).

In order to display a summary of the primary statistics of the distribution, a table with the cumulative distribution (cdf) (displaying the probability of a runout range less than the given quantile) , the corresponding quantile levels and their subsequent exceedance probability was made (fig.7-2). The table revealed only 5 per cent chance of exceeding 141 meters. When we know that the longest runout has a distance of 295 meters, which is more than twice this, it is evident that the distribution belongs to a light tailed and positively skewed pdf curve. The table for the quantiles of the runout distribution showed the 50^{th} percentile to be at 63 meters (fig.7-2), while the calculated sample proportion show that 81.1% of the samples have a range of 100 m or less. Furthermore, figure 7-1 along with figure 7-3 reveals that the density is highest on runout ranges between 20 and 40 meters.

Summary runout(S1)									
(CDF)	0.0	0.05	0.10	0.25	0.5	0.75	0.90	0.95	1.0
Quant. (x)	13.00	20.05	23.0	32.75	63.00	94.50	120.9	141.0	295
S(x)	1	0.95	0.90	0.75	0.5	0.25	0.10	0.05	0.0

Figure 7-2: Table of the cumulative distribution (cdf), quantile levels of runout and exceedance probability.

When examining the runout (S1) of the set we find that the distribution is positive or right skewed. That is, there are few higher values relative to lower values in our set, which ultimately causes a longer tail on the right. The skewness of the runout may be quantified and was found to be 1.85. In order to see if this is significantly different from zero we do a t-test and divide the estimated skew by its approximate standard error which is equal to 0.22. This gives the result of 8.36. The probability of obtaining a t-value of 8.36 by chance alone when the skew really is zero yielded a probability of $(6.56 * 10^{-14})$ which means that the null hypothesis of normality is rejected ($p<0.05$) and the distribution does show definite signs of non-normality. However, our sample size n=122, should by inference of the law of large numbers, be large enough to provide a reasonable estimate for our t-procedures.

The non-normality was also assessed by estimating the kurtosis of the distribution. In this respect a normal distribution will be bell shaped, whereas a kurtotic distribution deviates from a bell shape. Furthermore, a flat topped distribution is referred to as a platykurtotic and has a negative kurtosis value, while a pointed distribution is called leptokurtotic and has a positive value. If the distribution is perfectly normal the kurtosis value is equal to zero (Crawley 2005).

The kurtosis for the sampled runouts was calculated to be 4.85. In effect this means that the distribution exhibits a distinct leptokurtotic shape, which is significantly deviate from the attributes of a normal distribution.

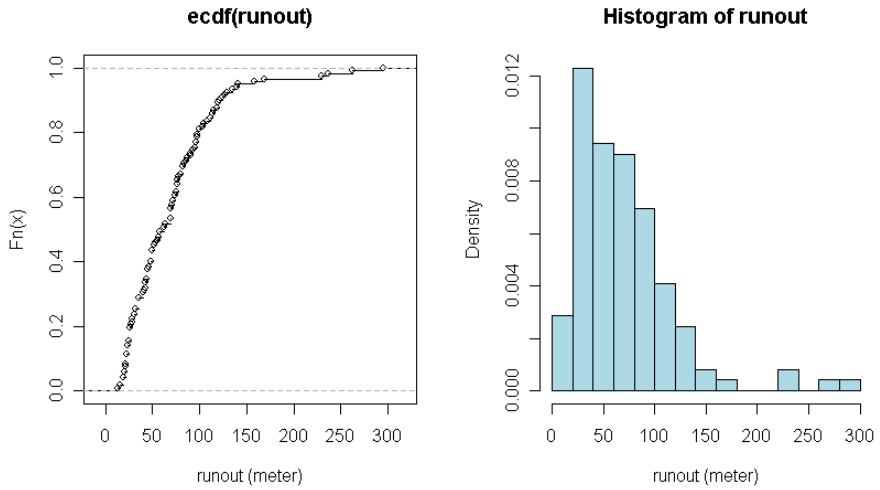


Figure 7-3: Graphical display of the cumulative distribution and a probabilistic density histogram of S1.

7.1.1 *Distribution types and density plots*

In many cases a curve can be drawn on the relative-frequency histogram, which is a very good approximation to the relative-frequency histogram. This is referred to as a density function, or density curve (Krumbein 1965).

The independent distribution of the runout beyond the talus (S1 in fig.2-1) was fitted to a lognormal, gamma and weibull type of distribution using the fitdistr() function in R. The fitting was done by means of maximum likelihood methods.

The intention of fitting distributions to the independent sample of runout was not only to find the best fitted density curve for the data, but also to conduct further investigations for the plausible distribution type. In this respect the values were compared in cumulative distribution plots and respective quantiles plots. In addition the estimated parameters for the fitted distribution may be used in a probability density function and a hazard function to the sampled distribution.

By comparing the theoretical distribution with the empirical distribution, it was found that the best fitted density curves belonged to a theoretical logarithmic distribution of the runout and a theoretical gamma distribution. The fitted distributions and the density curves were compared in a histogram, in addition to the original sample distribution and the density curves (fig.7-5, fig.7-7). Furthermore, the cumulative distribution of the theoretical fit was compared with the original distribution.

In order to see how the theoretical cumulative distribution deviated from the fitted empirical cumulative distribution, a Kolmogorov Smirnow-test was conducted for each "goodness of fit test". The results for this type of test were inconclusive as the p-value was 0.97 for both the fitted gamma distribution and the fitted lognormal distribution. For the weibull distribution the p-value was slightly lower at 0.80. Thus, the result suggests that the empirical cumulative distribution is closer to a lognormal, or gamma type of distribution.

In effect this means that all three distributions are statistically significant and can not be ruled out. Nevertheless, the fact that the null hypothesis was not rejected and that the discrepancy between the tested distributions is minor, indicate that the test may not be very robust for samples of this size (n=122).

In the case of the lognormal type of distribution (fig.7-5) the tail is longer and heavier relative to a gamma distribution (fig.7-7). As a consequence it would be an applicable type to distributions with low exceedance probabilities, as the density and length of extreme values may often be exaggerated compared to the original distribution. In this respect it may be seen as a conservative distribution type since the prolonged tail often leads to larger extreme values compared to other distribution types. By utilizing the MASS package and the fitdistr() function in R, the parameters of the distribution was acquired. The calculation produced a meanlog value of 4.04 and a standard logarithmic deviation of 0.66. These parameters was then transferred to a random distribution of the same length as the sample base, and served as the basis for further analysis.

In order to see how the observed (empirical) variable values compare with the theoretical quantiles of the distribution, a qq-plot was made. In addition a pp-plot was made to determine the how the theoretical cumulative distribution was related to the observed (empirical) cumulative distribution. In this respect it is important to bear in mind that the evaluated distributions, although having the same number of observations as our sample, are in fact based on random distributions. Subsequently the plot results are not absolute, but rather represent a general trend of the observed values relation with the theoretical values.
The result of the lognormal distribution showed that the relatively heavy distribution tail caused a deviation for the higher empirical values above 120 m in the qq-plot (fig.7-4), while the plot of the distribution function (pp-plot) showed an overall better correlation. Thus, the observed (fitted) random values (left of fig.7-5) are slightly heavier than the theoretical distribution curve (pdf) for lower exceedance probabilities.

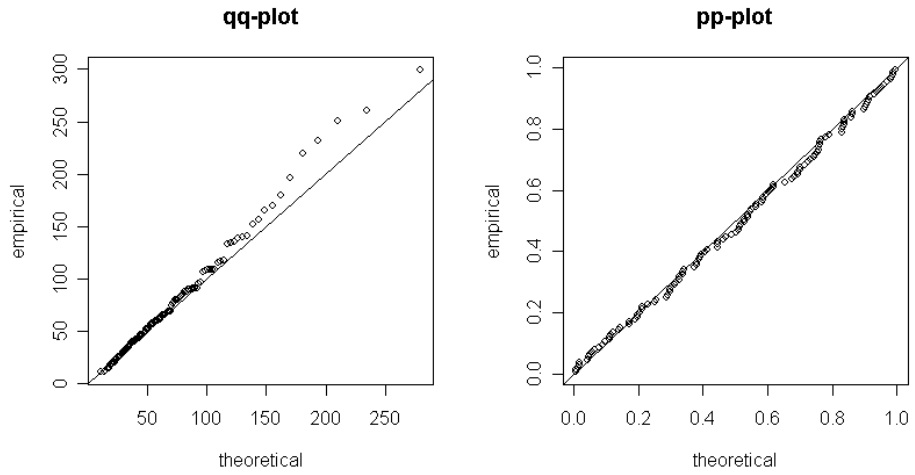


Figure 7-4: Plot of the qq-plot and pp-plot for the logarithmic distribution.

To further evaluate the distribution type, a histogram of the generated random lognormal values was plotted with a fitted probability density curve (pdf). The same pdf was added to a histogram of the sample basis for comparance (right of fig.7-5). The graphical display revealed that the distribution has significant similarities to that of a lognormal distribution, but is slightly deviate, particularly at values around the 50th percentile.

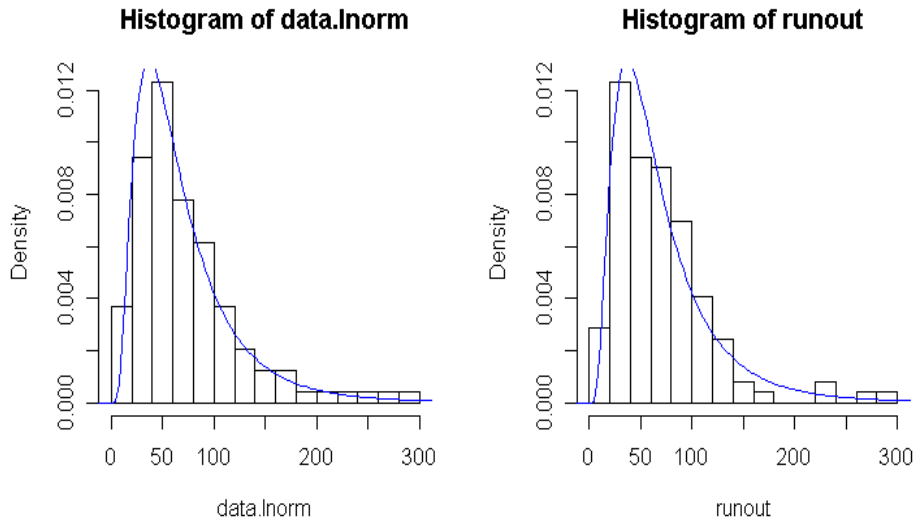


Figure 7-5: Histogram of the fitted random distribution with fitted density curve for random sample and databasis respectively.

The same fit assessments were carried out with respect to the fitted gamma distribution. The fitted gamma curve does not have the same tail length as the lognormal distribution (fig.7-7). However, the pdf curve is slightly broader and less kurtortic than the lognormal distribution. In effect this gives it a closer fit to the lower quantile values, while conversely the higher quantile levels are not as close to the curve. The relation between empirical and theoretical quantiles for the gamma distribution showed a slightly better fit on values between 125 and 150 meters when compared with the lognormal distribution (left of fig.7-6 vs. left of fig. 7-4).

That being said, comparison of the empirical and theoretical cumulative distribution (pp-plot) reveals a slightly better fit for the lognormal distribution (right of fig.7-6 vs. right of fig.7-4).

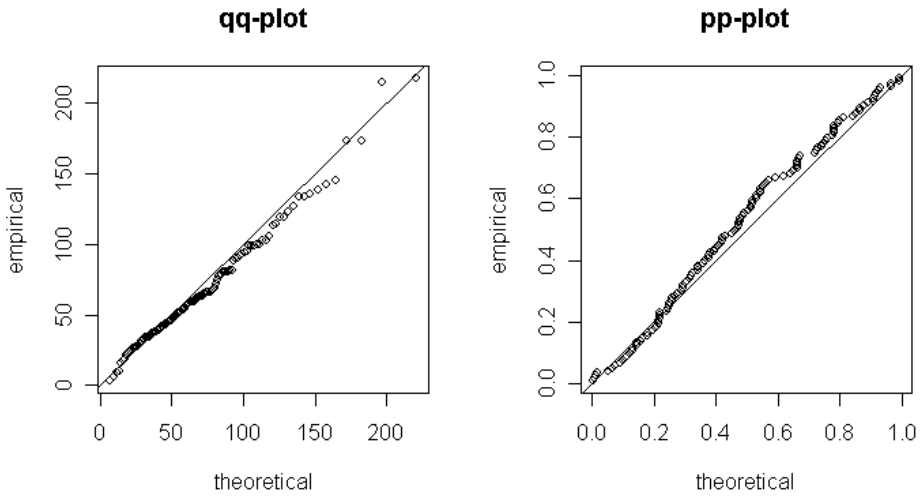


Figure 7-6: Plot of the qq-plot and pp-plot for the gamma distribution.

Ultimately the distribution test does not establish any conclusive distribution type for our population of rockfalls. However, we have strong reasons to believe that the population belongs to a significantly right skewed distribution type as opposed to a more bell shaped or normal distribution type.

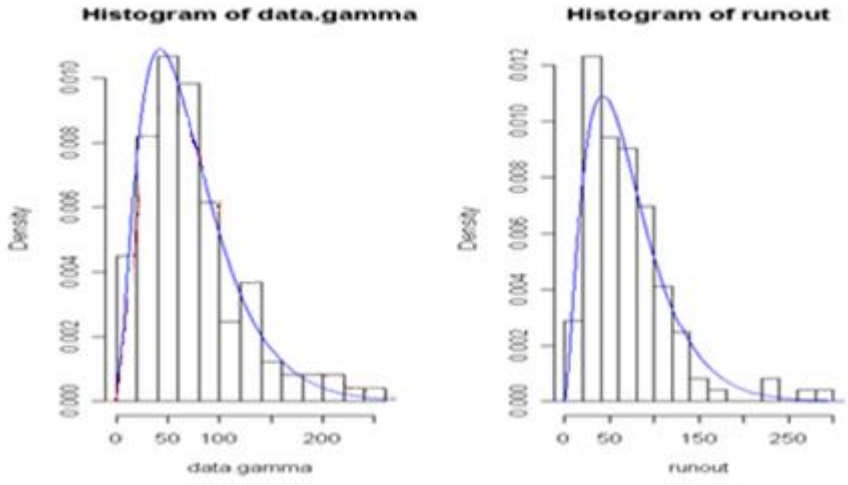


Figure 7-7: Graphical display of the gamma distribution with the random based sample and the databasis.

7.2 Relation between runout distance and total height

Before commencing the model analysis, it is reasonable to believe that the height from the source point of the rock down to the talus slope would be an important variable for the range of the boulder. The reason for this is ultimately founded in the potential energy (mgh) that the rock possesses while being attached to a cliff. On the other hand, the subsequent energy loss at impact with the talus is likely to increase with the height of the cliff (Domaas 1994).

A simple linear regression for the runout beyond the talus (S1 in fig. 2-1) as a response of the total height was performed in R. The regression gave a squared correlation of 0.542.
When transformed to a log-log plot the squared correlation was increased to 0.70. However, the original distribution of runout as a response of the total height (htot) did not show signs of having a non-linear relation. Subsequently a calculation by means of the corresponding derived power function did not lead to an increase in the correlation value.

Except for the longest runout values, the resultant plot does not display a significant variance. However, these observations (that is observation nr. 122,121,120 and 116) contribute to a significant pull on the regression line. This is particularly notable on the original plot (fig. 7-8), where the high leverage of the four observations effectively causes the regression line to drop from the main distribution pattern. Subsequently, the longest runouts beyond 200 meters seem to belong to a different distribution. In larger Mohr-Coulomb driven mass flows we see the same phenomena with respect to the longest runout distances. Rockfalls however, is restricted to individual fragments which imply a much smaller volume. Subsequently, they are significantly different from a rockslide. The subjected body is restricted to individual fragments of a much smaller volume, and thus the contact with the surface is not as extensive as for a dry mass slide. In the end it is reasonable to believe that this is due to the large γ value that these observations have.

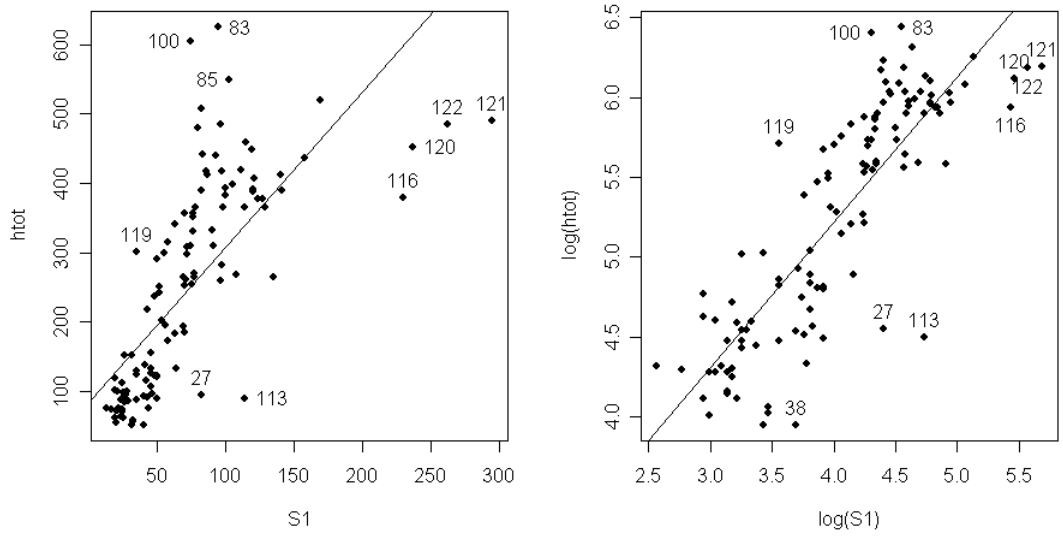


Figure 7-8: Plot of (htot) vs. the runout beyond the talus (left), and the logarithm of the same variables.

7.3 β-Ψ model

The β angle has formerly been used by Albert Heim as a means for predicting runout ranges of so called "sturzstroms". It is defined by a straight line between the runout range of the boulder and the uppermost part of the cliff. According to Heim the line may be used as an estimated energy line which expresses the rate of frictional dissipation of energy. In this respect an average friction value of 31°, which corresponds to a friction coefficient of 0.6 has been utilized in the program Rockfor.net to calculate rockfall hazard in a forested slope.

In a strictly topographic sense we would expect the angle to lower with the increased runout distance from the cliff. Empirically this would mean that the longest runouts have the lowest β angles. A simple scatter plot between the β angle and the runout reveals that the hypothesis is somewhat oversimplified (fig.7-9).

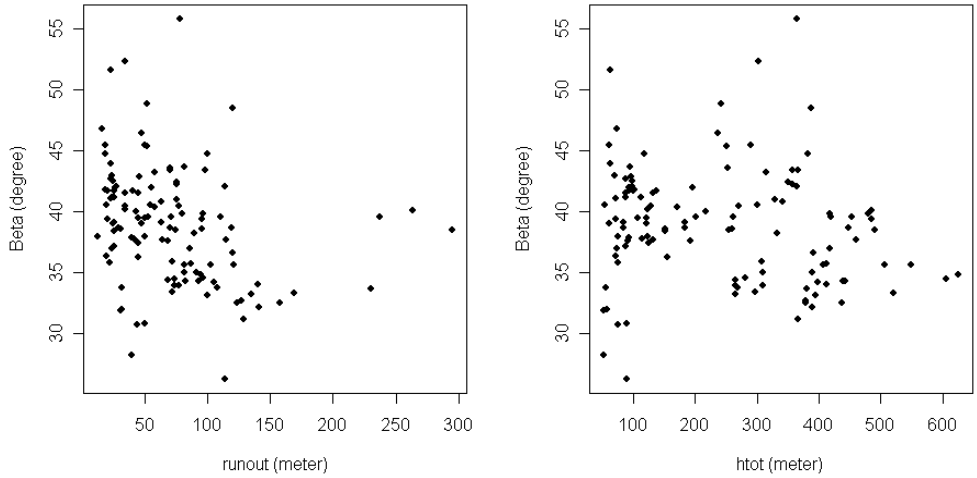


Figure 7-9: Plot of the relation between (β) and S1 (left). Plot of β and the total height (htot) (right).

That being said, the plot shows that none of extreme runouts are below 40.1°. Furthermore, the comparatively longer runouts between 125 and 169 meters are skewed towards lower β angles between 30 and 35 degrees.

In the plot displaying the angle's relation to the total height (on the right), a scale effect is observable. In this respect it suggests that the observations under 100 meters belong to another distribution. Thus, the plots reveal that the variance of β angles is particularly large at relatively smaller heights and subsequently shorter runout distances.

The ψ angle is described as the straight line between the talus terminus and the top of the cliff. In a topographic sense it describes the length of the scree accumulation and its relation to the top of the cliff. Empirically we would expect the ψ angle to lower in accordance with a more inclined cliff or a longer or more mature talus (S2, H2). According to Kotarba's field studies in section 4.2.5 there exists a topographic relation between the inclination of the cliff (θ) and the profile of the talus.

To further investigate this, a scatter plot between θ and ψ was made. The plot (fig.7-10) revealed a definite correlation between the two variables, as evident by a squared correlation value of 0.59 and a residual model error of 6.47 degrees. However, the variables exhibit a large spread and there is no overall tendency leaning towards over or under prediction. Instead the general trend is that of an increased variance with increased angles. Thus, the plot exhibits a significant degree of heterodasticity.

The most deviate observations identified did not show any significant connection with respect to the runout beyond the talus. Ultimately the result suggests that cliffs (H3) which are comparatively more vertical are more plausible to have a greater residual error. Referring again to Kotarba's field studies, we now that these cliff types are more likely to have steeper and straighter talus deposits than more inclined cliffs. Furthermore, we see that the steepest cliffs with respect to the ψ angle leans towards relatively higher H3 intervals.

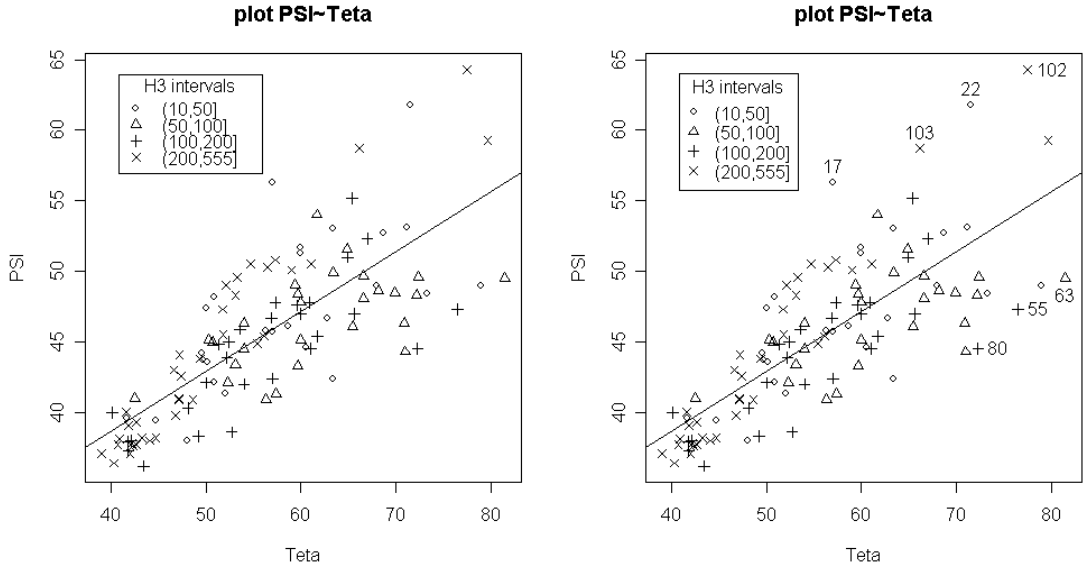


Figure 7-10: Plot displaying the relation between ψ (talus angle) and θ (inclination of the cliff).

The simple linear regression that yielded the highest correlation between two empirical parameters was found between the β angle (fahrböschung angle) and the Ψ angle (talus angle). A regression analysis with β as a response variable of ψ gave an adjusted squared correlation value of 0.80 and a standard error of 1.62 for the intercept. Thus, the result indicates that there is definite relationship between the two parameters. The model performance was verified by the by the ANOVA F-test which produced a p-value of $2*10^{-16}$ for the null hypothesis. In addition the graphical display of the distributions (Appendix D) confirms this. The analysis produced the following regression equation:

$$\beta = 3.926 + 0.768\psi$$

The diagnostic plots of the regression (Appendix B) include residual plots, standardized residuals, a qq-plot with respect to the observed (empirical) standardized residuals and the theoretical quantiles and lastly a leverage plot.

The resulting regression plot (fig.7-12) and diagnostics revealed that observation nr.113, which occurs in the lower end of the spectrum, is a leverage outlier, and ultimately has a high influence on the least squares regression line. In this respect its influence pulls the intercept of the line down. In order to make the general influence results more prominent, a histogram (fig.7-11) displaying the Cook's distance of each observation was made. The estimated result suggests that particularly point 113, 102 and 17 are influential points in the regression analysis. However, the reader should be aware of the fact that the Cook's distance method to determine influence, while displaying the influence very clearly is not as sensitive as standardized difference in fit method (DFFITS).

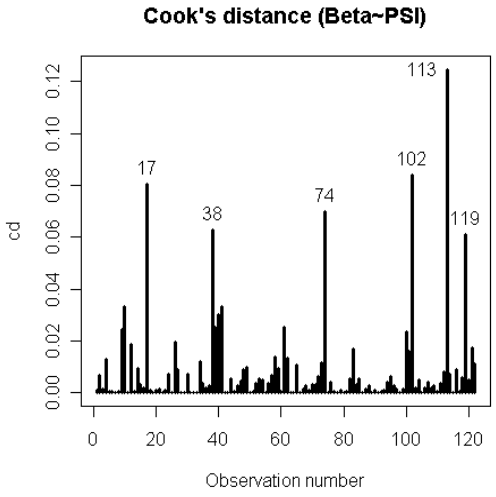


Figure 7-11: Cook's distance histogram of influential observations in ψ-β relation.

A closer inspection on observation 113 reveals that it has a relatively low total height potential of 114 m, which is not much above the 25 per cent htot-quantile of 95.25 m. Even so it has a runout (S1) that is above the 75 per cent quantile for the independent distribution. However, the data also reveal that it has a γ angle which ranks among the highest with a value well above the 15^{o}, which is thought to promote particularly long runouts. In this respect the observation is an important indication that the angle of the topography beyond the talus has a significant influence on the length of the runout.

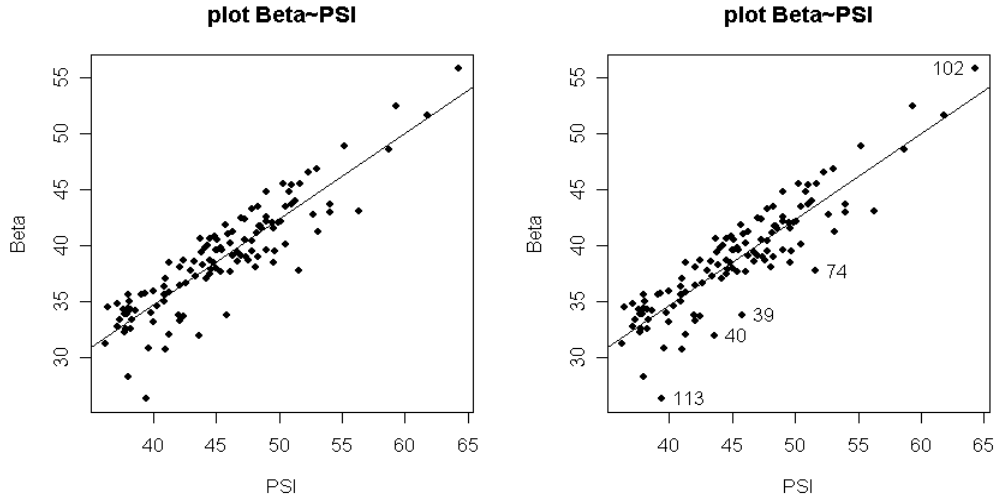


Figure 7-12: Simple linear regression result between ψ and β angles, with obs. identification (left).

A conversion of the predicted regression values from degrees to corresponding predicted L1 values was carried out. This was achieved by application of simple trigonometrical calculations of the total height (htot), with respect to the predicted β and ψ values. The converted L1 values was then calculated to predicted S1 values by utilizing a constant γ value

equal to the distribution mean. As a direct consequence of this, a small part of the over and under prediction may be a result of a constant γ value used during the conversion process. The conversion from predicted angles to new lengths was carried out utilizing the following formula:

$$(\boldsymbol{S1}+\boldsymbol{S2})_{\boldsymbol{pred.}} = \frac{L_{tot}-L_{2+3}}{\cos(\gamma_{mean})} + \boldsymbol{S2} = \left(\frac{\frac{Htot}{\tan(pred.\beta)} - \frac{H2+H3}{\tan(pred.\psi)}}{\cos(\gamma_{mean})}\right) + \boldsymbol{S2}$$

The resultant plotted values (fig.7-13) consist of the predicted (S1+S2) values against the original (S1+S2) values. This produced an increased adjusted square correlation value of 0.88 for the whole dataset and a standard error on the intercept of 8.7. The model's standard error (the residual standard error) was 50.8 meter.

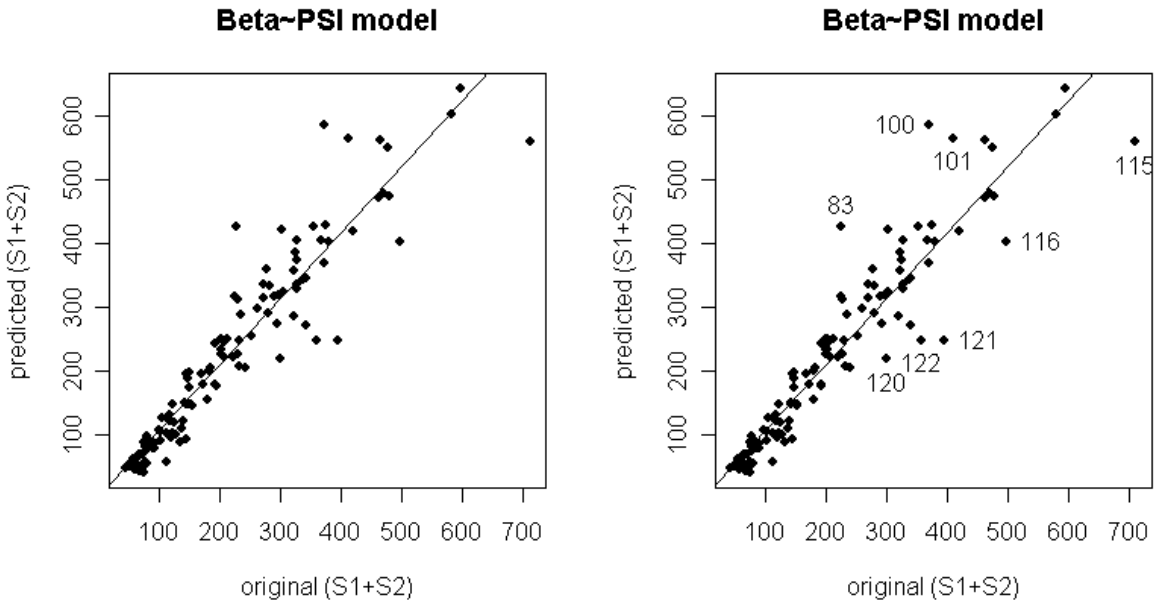


Figure 7-13: Plot displaying the resultant β-ψ model.

The result shows that the model has a prominent under prediction for five of the observations. This includes the four extreme values of S1 beyond 200 meters, in addition to the fourth longest observation of 169 meters. The latter is the most notable deviate observation. It is a significant high leverage outlier, posing the longest (S1+S2) value of the original sample. Subsequently, its contribution pulls the regression line down, while decreasing the squared correlation value.

In order to compare the model performance with the original data a non-parametric bootstrap was performed with 1000 resamples of both the original sample (fig.7-14) and the model (fig.7-15). For values at higher cumulative levels the method utilized in R revealed some signs of instability.

Subsequently, particularly the confidence intervals for extreme values, that is around the 95th percentile level, should be taken as more of a rough estimate than an absolute interval. The tables reveal that the model underpredicts values between the 10th and 25th percentile, while it overpredicts at a quantile level of 75.

For the longest runouts, that is around the 95th percentile, the model tends to underpredict. Furthermore, the descriptive statistics (fig. 7-16) show that the model has a significantly greater spread than the sample basis. Overall it can be said that the model shows a slight heterodastic tendency, which is increased variance with larger (S1+S2) values. In effect this means that the distribution of the predicted runout distances has a slightly broader and less kurtotic distribution curve than the sample basis.

(S1+S2)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(71.98, 73.24)	(101.5,104.0)	(192.2, 193.0)	(301.2, 303.6)	(469.5, 475.5)
CI=0.25	(71.78, 74.50)	(94.7, 104.6)	(188.1, 201.5)	(299.6, 306.0)	(465.3, 475.5)
CI=0.5	(69.98, 79.01)	(89.7, 114.0)	(184.2, 205.4)	(281.1, 313.2)	(462.5, 478.3)
CI=0.75	(68.69, 80.08)	(86.5, 117.7)	(182.7, 216.5)	(277.4, 323.9)	(460.6, 521.3)
CI=0.95	(65.62, 84.98)	(78.2, 124.7)	(164.7, 236.3)	(275.9, 344.6)	(363.1, 545.6)

Figure 7-14: Table of quantiles (p) and confidence levels (CI) for the original sample of runout (S1+S2).

(S1+S2P)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(60.80, 61.00)	(91.4, 92.3)	(194.5, 201.4)	(320.0, 322.7)	(478.4, 478.4)
CI=0.25	(60.80, 65.60)	(88.7, 95.1)	(193.9, 203.3)	(316.7, 323.5)	(405.8, 483.5)
CI=0.5	(60.20, 66.60)	(83.6, 97.4)	(181.5, 205.2)	(305.9, 325.3)	(405.8, 485.1)
CI=0.75	(53.10, 67.30)	(80.1, 99.2)	(177.2, 222.1)	(304.6, 341.3)	(392.8, 528.6)
CI=0.95	(43.30, 72.40)	(65.0, 105.7)	(155.8, 249.5)	(282.5, 368.4)	(369.4, 536.1)

Figure 7-15: Table of quantiles (p) and confidence levels (CI) for the predicted sample of runout (S1).

Desc. Stats.	st. dev.(s)	variance	mean	median
Sample	134.59	18115.06	215.12	192.8
Model	149.14	22243.39	225.42	202

Figure 7-16: Table of the descriptive statistics of the model and the original sample.

7.4 α – δ model

The empirical relation between the talus and the shadow angle (α) as a predictor for runout of rockfall has formerly been commented on by Evans and Hungr. The angle is the straight line between the highest point of the talus and the end point of the longest runout boulder of the rockfall (fig.2-1, Appendix A). Investigations by Evans and Hungr reported a minimum shadow angle of 27.5^{o} based on 16 talus slopes in British Columbia, while other comparable studies has reported an interval between 22 and 30 degrees (Dorren 2003).

The independent distribution of the α angles show a significant spread from 15.8^{o} to 39.3^{o} with a standard deviation of 3.91 degrees. This suggests that earlier reports of a minimum shadow angle are somewhat oversimplified, or based on too few observations. Furthermore, an interval sorting of the distribution reveal that 85 of the observations are between 22^{o} and 30^{o}, 25 are between the minimum of 15.8^{o} and 22^{o}, while lastly 11 are between 30^{o} and the maximum of 39.3^{o}. Thus, close to 30 per cent of the distribution are outside of the projected interval between 22 and 30 degrees.

In a strictly geomorphologic sense, we would expect the maturity of the talus to have an influence on the shadow angle. This theory was tested by means of a scatter plot between the shadow angle and the talus height (H2). The result (fig. 7-17) revealed that the comparatively shorter runouts had the largest angles. This is explainable by the fact that a relatively shorter runout boulder is closer to the talus and would ultimately need a steeper angle to reach the top of the scree. Even so, as a result of this, we would still expect the talus angle (δ) to be steeper than it is for the longest runouts. The plot on the right of figure 7-17 confirms this, as a large amount of the observations with relatively larger α-angles also possess relatively larger δ angles. Subsequently, the figure suggests that poorly evolved talus slopes clearly tend toward higher angles when compared to those of a more mature talus. This is what we would expect since a more evolved talus which has been predisposed to rockfall, ultimately would have more gravity driven “fall sorting substrate” at the end of the toe, which in turn would lead to a less inclined talus slope over time.

Furthermore, we would be inclined to believe that the normal component of weight would be increased in comparatively steeper slopes. As a consequence of this, it would be fair to assume that the retardation caused by friction with the substrate increases with the inclination of the talus slope (Domaas 1985).

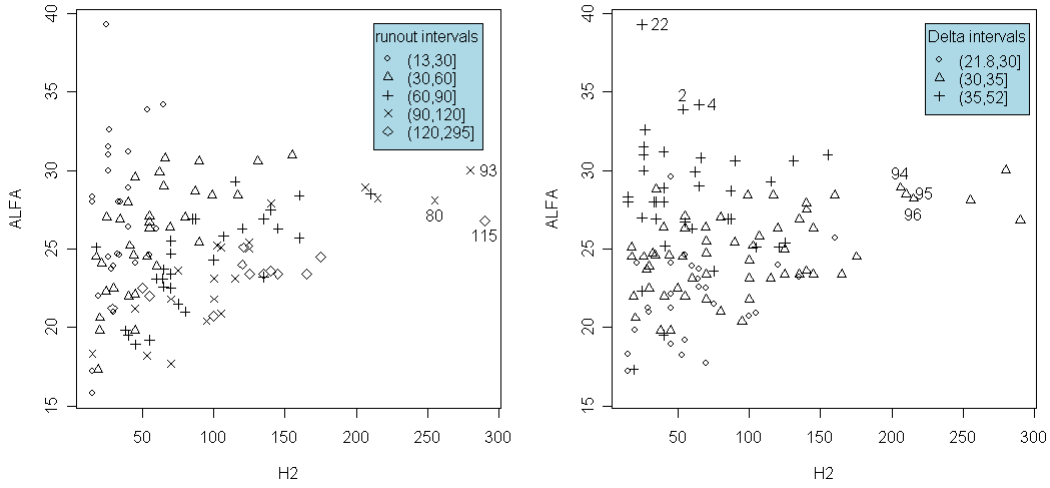


Figure 7-17: Plot displaying the shadow angle (α) in degrees with respect to the height of the talus (H2).

Plotted together, the α angle (the shadow angle) and the δ angle (the inclination of the talus) yielded an interesting linear relation (fig.7-19). A simple linear regression analysis with α as a predictor for δ produced an estimated adjusted square correlation value of 0.54, with an intercept standard error of 1.83. The corresponding model residual error was 2.67 degrees. Furthermore, the analysis produced the following linear relation between the two parameters:

$$\alpha = 3.60 + 0.65\delta$$

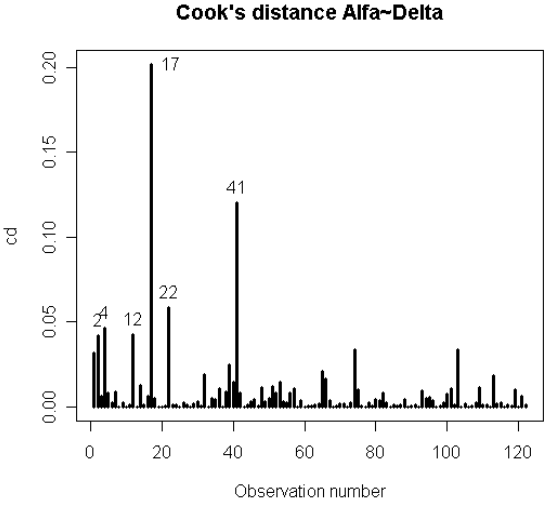


Figure 7-18: Cook's distance histogram displaying influential observations in the α-δ relation.

The diagnostic plots of the regression (Appendix C) reveal that the residual error is relatively constant throughout. Furthermore, the normal qq plot between the standardized residuals of the regression model and the theoretical quantiles display significant signs of normality. This is also confirmed by looking at the distributions in Appendix E. For the α-δ relation, the most influential point of the regression model is difficult to determine simply by looking at the plot. However, observation 22 is likely to have a high influence on the regression line. This assumption is based on the high leverage of that observation which effectively pulls up the regression line. The observation is plausible to come from a site with a small immature talus as the runout of the boulder is no more than 23 meters which is the same as the 10th percentile of the runout distribution. Furthermore, we see that observation 41 in the other end of the plot has the potential of being a high influence point. By reference to Cook's distance (fig.7-18), these two points together with largest influence observation (17) comprise the most deviate points

in the model. These have runout distances beyond the talus of 24 and 25 meters respectively, while observation 41 has a runout of 41 meters.

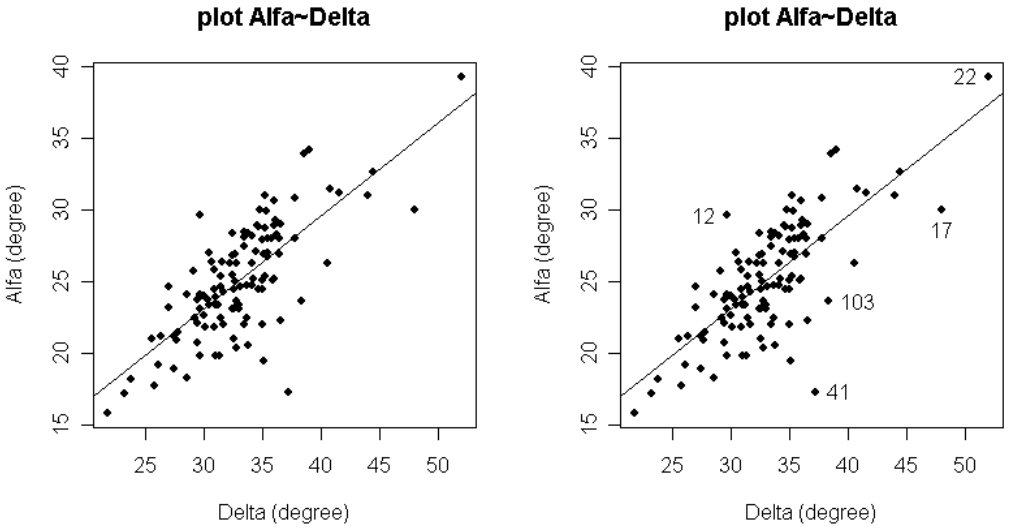


Figure 7-19: Plot of the regression analysis between the α and δ parameter.

A conversion using the same methodology as for the ψ-β relation was carried out, that is utilizing trigonometrically methods and a constant distribution mean of the γ angle corresponding to 9.0 degrees.

For the α-δ model, the prediction function became the following:

$$(\boldsymbol{S1}+\boldsymbol{S2})_{\boldsymbol{pred.}}=\frac{L_{1+2}-L_2}{\cos(\gamma_{mean})}+\boldsymbol{S2}=\left(\frac{\frac{H1+H2}{\tan(pred.\alpha)}-\frac{H2}{\tan(pred.\delta)}}{\cos(\gamma_{mean})}\right)+\boldsymbol{S2}$$

The conversion lead to a significant increase in the model performance as the regression analysis yielded a value of 0.96 for the converted lengths (S1+S2). The reason for this is in part thought to be an effect of the relatively shorter distance (L1+L2) compared to the ψ-β analysis where the whole length (Ltot) was derived in basis of the predicted β value. In addition further investigations on the residuals revealed that the most deviate points also are the one with the comparatively shorter runout distances. Subsequently the error becomes smaller when S2 is added in the regression analysis. This suggest that the model would be more suitable at mountainsides with relatively longer runout distances beyond the talus slope (S1), that is at locations which are likely to have longer and more evolved talus (S2).

From the plot (fig.7-20) we see that the deviation of observations does not seem to increase with lower exceedance probabilities. The models standard error (residual error) was 32.5 meters, while the standard error of the intercept was 4.55. The five identified observations in fig.7-20 all correspond to runout ranges (S1) above 80 meters. The longest of these are observation 115 and 121, with distances of 169 and 295 meters respectively.

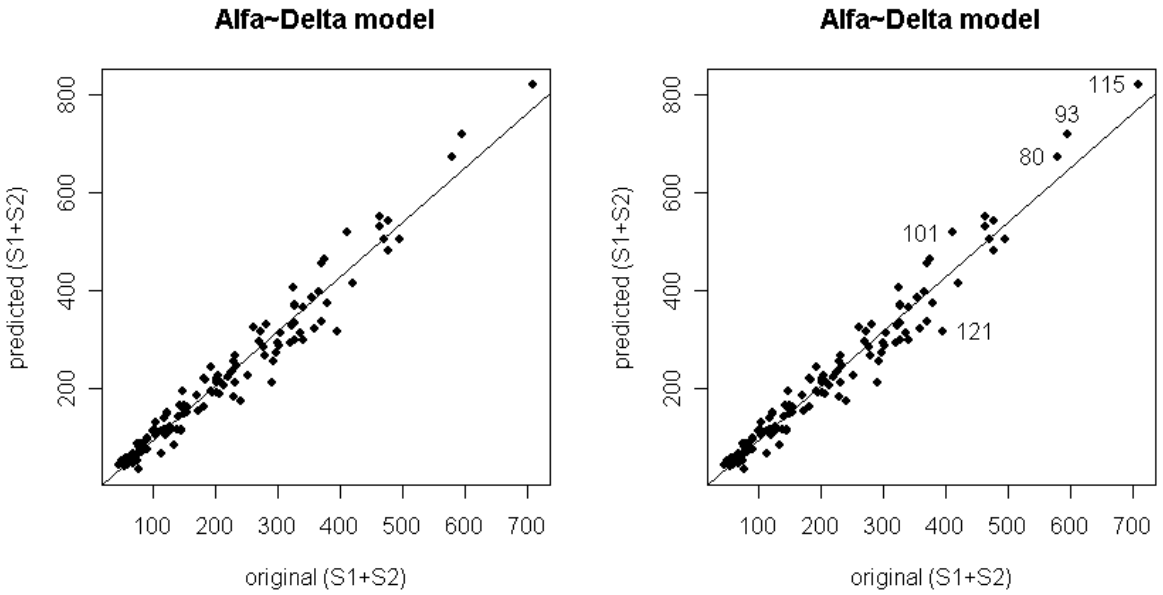


Figure 7-20: Plot of the original runout (S1+S2) against the predicted runout (S1+S2).

In order to compare the model with the original distribution at different levels of runout distances, a non parametric bootstrap with 1000 replicates was performed. The tables of bootstrapped runout ranges (fig.7-21, fig.7-22) reveal that the model significantly underpredicts observations around the 10th percentile, while it tends to overpredict at the 95th percentile. However, the model shows a moderately better accuracy between the 25th and 50th percentile.

The descriptive statistics (fig.7-23) reveal that the spread is significantly larger as evident by a larger variance. In addition the mean is slightly higher, indicating that the predicted distribution has a somewhat higher frequency of comparatively longer runouts. The median is very similar, which means that the distribution variance has a net over and underprediction which is close to 0.

(S1+S2)	**p=0.1**	**p=0.25**	**p=0.5**	**p=0.75**	**p=0.95**
CI=0.1	(71.98, 73.24)	(101.5,104.0)	(192.2, 193.0)	(301.2, 303.6)	(469.5, 475.5)
CI=0.25	(71.78, 74.50)	(94.7, 104.6)	(188.1, 201.5)	(299.6, 306.0)	(465.3, 475.5)
CI=0.5	(69.98, 79.01)	(89.7, 114.0)	(184.2, 205.4)	(281.1, 313.2)	(462.5, 478.3)
CI=0.75	(68.69, 80.08)	(86.5, 117.7)	(182.7, 216.5)	(277.4, 323.9)	(460.6, 521.3)
CI=0.95	(65.62, 84.98)	(78.2, 124.7)	(164.7, 236.3)	(275.9, 344.6)	(363.1, 545.6)

Figure 7-21: Table of original runout (S1+S2) values bootstrapped with 1000 replicates.

(S1+S2P)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(63.68, 64.58)	(102.5, 106.7)	(189.9, 192.9)	(289.4, 298.9)	(517.2, 518.5)
CI=0.25	(63.64, 69.80)	(100.5, 108.9)	(188.5, 195.5)	(285.7, 300.7)	(504.2, 529.8)
CI=0.5	(61.68, 72.45)	(94.7, 116.1)	(178.2, 203.2)	(282.3, 306.1)	(502.8, 532.2)
CI=0.75	(55.96, 72.98)	(93.2, 121.7)	(170.3, 219.0)	(276.2, 318.0)	(492.3, 571.5)
CI=0.95	(47.23,77.41)	(83.8, 124.1)	(163.4, 225.3)	(266.7, 343.3)	(366.7, 619.8)

Figure 7-22: Table of (S1+S2) values for α-δ model bootstrapped with 1000 replicates.

Desc. Stats.	st. dev.(s)	variance	mean	median
Sample	134.59	18115.06	215.12	192.8
Model	153.50	23562.42	221.52	191.3

Figure 7-23: Descriptive statistics of model against original distribution of (S1+S2) values.

7.5 Multiple regression - preliminary computations

The multiple regression tests were conducted with the simple regression results in mind. In this respect it was believed that the β and ψ parameters would be important factors for the response. The intention of the preliminary preparation was to detect relationships between the selected variables of the analysis. That is, to check which explanatory variable that would be the most crucial for the response variable, which in this case is the runout (s1). In order to not overcomplicate matters, the computations were initiated with six different explanatory variables. These were the total height of the mountainside (htot), β, α, ψ, γ and δ.

To show the relationship of all the selected variables in the model, a paired matrice (fig.7-24) was made that displayed all correlation possibilities. The matrice revealed a substantial amount of scatter between the different explanatory variables. Furthermore, other than a positive relationship with the total height (htot), it did not reveal any clear relationship between the response and the different independent variables. Subsequently the matrice raises the question of whether the interaction processes of the independent variables may be significant for the outcome of the response.

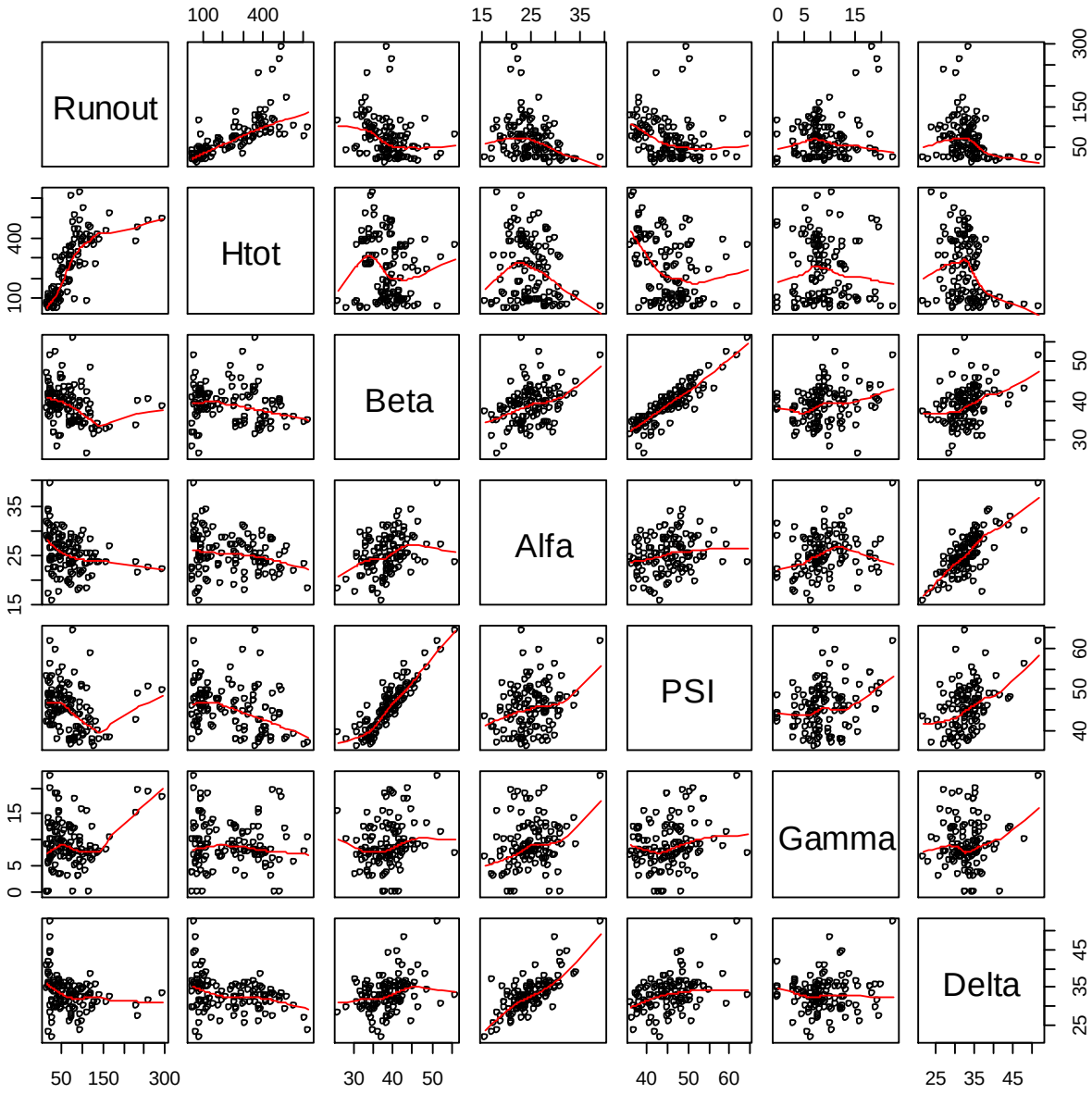


Figure 7-24: A paired matrice of the response (runout) and the different independent variables.

In addition to the paired matrice a non parametric method referred to as a generalized additive model (GAM) was conducted on the data (fig.7-25). The test is blending properties of multiple regressions with additive models. This is done by replacing the independent parameter terms $\beta_n x_n$ with functions $f(x)_n$. The functions are random, thus providing better fits to the data than other methods (Hastie 1990). The test displayed narrow confidence intervals with the total height and the gamma angle, thus confirming a close relationship with these parameters. However, the other angles have a larger deviation in the confidence interval for larger values of the angle.

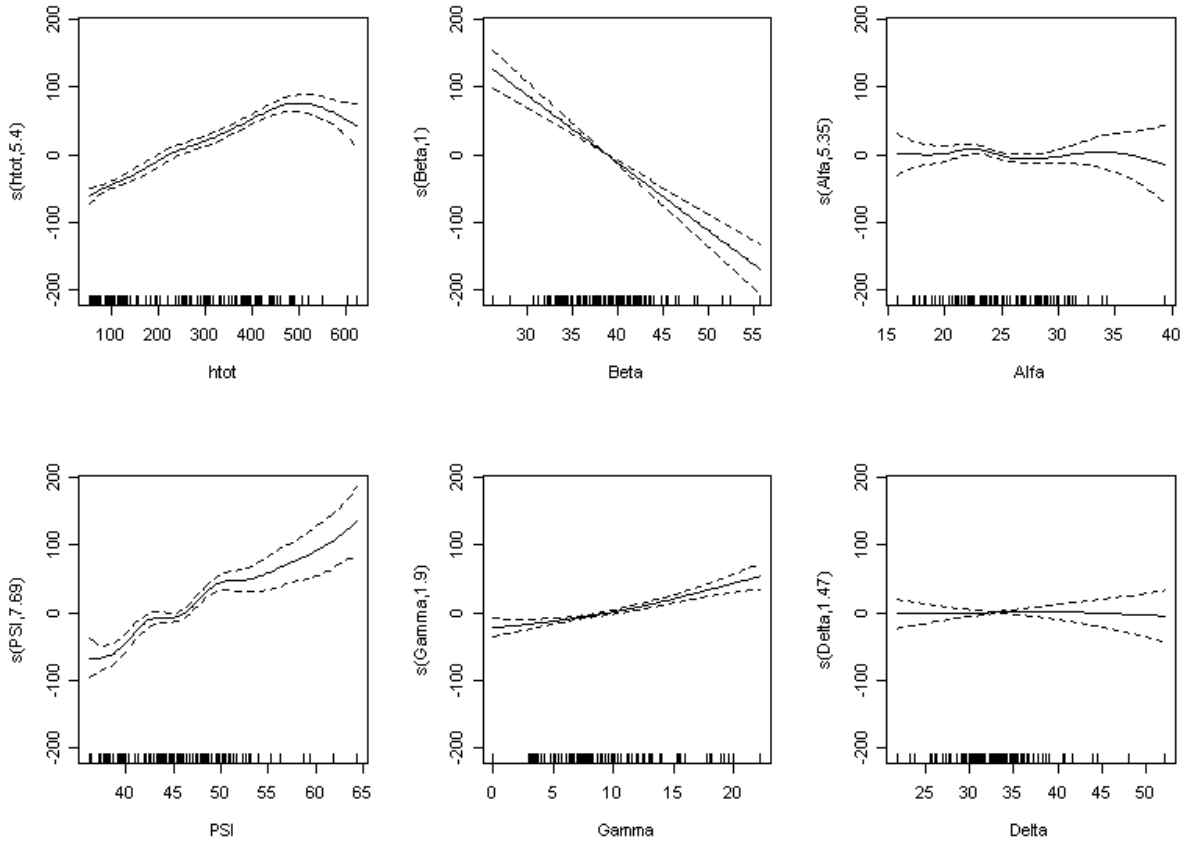


Figure 7-25: Figure displaying a gam of the different independent variables.

In order to determine whether the regression model shows evidence of complex interactions, a regression tree was constructed. The regression tree is based on a binary recursive process. This is an iterative process which splits the data into partions and then splits it further up into branches (refer to section 5.4.4). The tree exhibits a hierarchical structure with threshold values for each explanatory variable. In the right part of the tree we have the higher mean runout, while in the left part we have the lower mean runout. At the end of the branches we find the mean runout levels. Furthermore, the length of the branches is also important since the longer the branches of the tree, the more deviance is explained (Crawley 2005).

The result displayed in (fig. 7-26) shows that the most important variable affecting runout is the total height. This variable has a mean threshold value of 365 meters, meaning that the longest mean runouts occur for total heights above this threshold. The second most important variable is the γ angle. For the γ angle the threshold value is set at 14 degrees, which provide a higher mean runout of 225 meters. For high total potentials the β angle is significant at lower gamma values ($<13.95^{o}$).

If we look closer at the variables most important for lower potentials (the left main branch), we see that the height itself is the most important variable posing a threshold of 163.5 meters. Moving further down at the levels of lower potential, we find that a higher β angle is significant for higher mean potential levels.

In effect this means that the β-angle at low potentials (left main branch) is associated with higher potential (>163.5), while at higher potential (right main branch) it is associated with lower gamma angles (<13.95^{o}).

As a consequence it is associated with longer mean runout distances for low potentials, and shorter mean runout distances for high potentials. This fact, in addition to the fact that different independent variables are significant for high and low mean levels of runouts, indicates that the interaction structure between the data may well be of a complex nature. It is also interesting to note that a high mean gamma angle is relevant for longer runout ranges, while the total height is the most crucial parameter for lower mean runouts. The test confirms earlier notions that the gamma angle is one of the most important parameters (that is second most important) for longer runout distances.

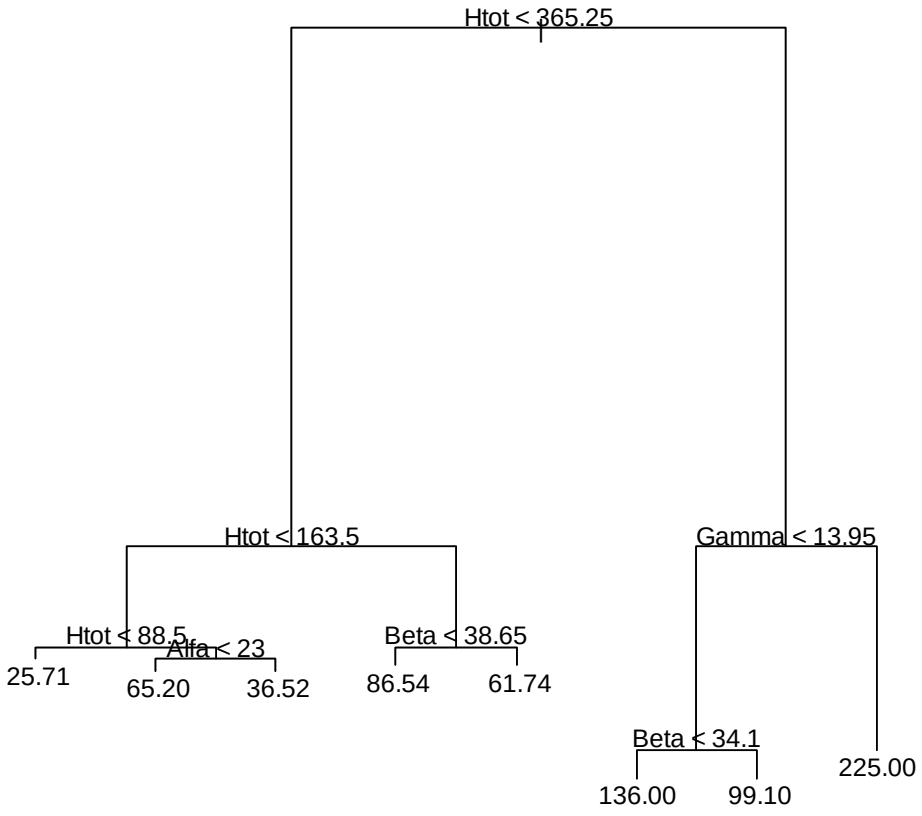


Figure 7-26: Regression tree displaying the utilized explanatory variables in the model.

7.5.1 *Multiple regression-evaluation of model*

With the preliminary computations in mind, the multiple regression analysis was initiated. In order to not overlook any significant terms the regression analysis was commenced with the most complicated model (Appendix F). The first model consisted of all the exploratory variables and its squared counterpart. The result revealed that the total height and its respective squared explanatory variable denoted by I(Htot^2), (where I is a function to assess squared parameters in R), is significant. In addition the squared term of the γ-angle is an important factor in the model. The squared correlation produced a value of 0.84 and a model probability of $2.2*10^{-16}$.

With the intention of establishing the relevance of interaction between the variables, three models with five different two way interactions were carried out. The results of these three models indicated that the principal significance was acquired in the interaction terms between the total height and the different angles. In this respect the most significant term was achieved between the interaction terms (htot:Beta) and (htot:PSI). Definite significance was also found in the interactions between the γ angle and the height.

With the intention of model simplification, the most significant parameters from each model were plotted together in new models. The least significant terms would then be left out, while the significant terms would be passed on for the next model.

The first model resulted in a relatively low significance of the β variable. The reason for this was thought to be a multicollinearity effect between the β and ψ angles, that is a relatively high correlation between the variables that in effect impairs the model. As a consequence of this suspected multicollinearity and subsequent low significance, the β-angle was left out for the next analysis. In addition an interaction term between the δ angle and the squared height was substituted into the model. This parameter substitution did not produce any significant difference in the model performance. Subsequently the δ parameter was left out from the next model.

By now the analysis had indicated that there is a definite interaction between the height and the β and ψ angles, and possibly also with the γ angle. Thus combinations of these three angles in addition to the height were utilized for a final attempted model, which also included a three way interaction (Htot:PSI:Gamma). This managed to improve the model performance slightly by evidence of a lower residual standard error of 9.27 meter and a minor increase in the adjusted squared correlation to 0.97 (Appendix G). As a consequence the end model used for further statistical analysis was the following:

Runout ~ htot + PSI + I(Gamma^2) + htot:Beta + htot:PSI + htot:PSI:Gamma

We note that the explanatory variables consist of the total height, the talus angle (ψ), the squared γ angle, and sets of three different interaction variables.

The latter variables suggest that there is a combined effect between the height and the different angles at different levels. Thus, the effect of one explanatory variable (e.g. htot) is dependent on the particular value (level) of the assessed angle.

In order to display the relation in two-way interactions with runout, we can make a 3D plot with the runout as the response (vertical z-axis), while the height and angle is displayed as the horizontal x and y axis respectively (Appendix H, Appendix I). The 3D-plot in Appendix H shows a least squares plane and the relation of ψ and htot as explanatory variables for the response of runout beyond the talus. The figure is in many ways indicative of the strong dependence of the total height for the runout range. Furthermore, we note that the longest runouts seem to occur at the relatively lower levels of ψ. Thus, there seems to be a combined effect between height and relatively low ψ values which increases the runout response. However, as these isolated constitute only one out of six explanatory variables of the runout response, the interaction between the variables is not ideal. The same type of 3D plot was made for the relation between the β angle and the total height (Appendix I). Once again we see that the longest runouts occur at the lower scale of β values.

The predicted response from the multiple regression models was plotted with the original sample of runouts beyond the talus (fig.7-27). The model produced an adjusted squared correlation of 0.97 and a residual standard error of 9.4 meter. Furthermore the model performance denoted by the Anova F-statistic yielded a p-value of $2.2*10^{-16}$. Thus, there is a clear significance in the linear regression which effectively disproves the null hypothesis. Interestingly the model seems to capture the distributional extreme values well. This suggests that the significance of the γ angle at different potentials is well incorporated within the model frame. In this respect, it is particularly encouraging to note that two of the low exceedance probability runouts falls within the 95 per cent confidence band for the mean response.

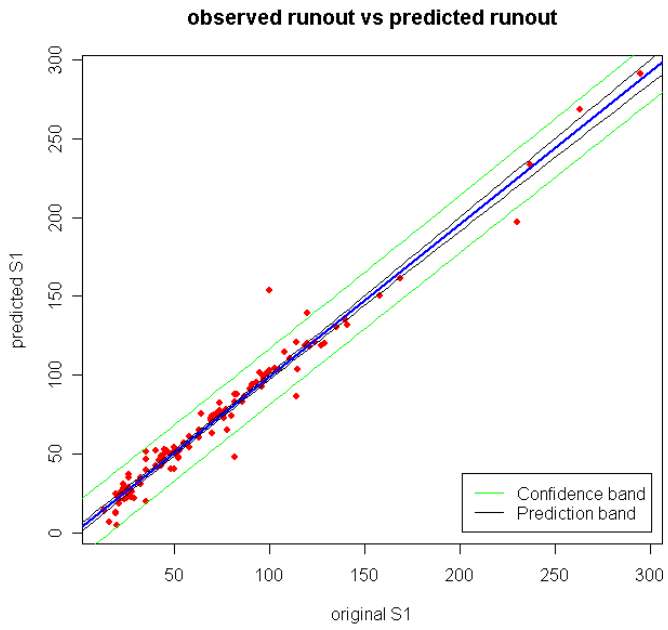


Figure 7-27: Plot of runout (S1) vs. predicted runout with 95% confidence and 95% prediction band respectively.

The model performance may be assessed by utilizing non parametric methods. This was conducted by comparing the quantiles and the bootstrapped values of the model with that of the original sample. From figure 7-28 and 7-29 displaying the bootstrapped values, we see that the distribution of the model is closely related to the sample basis.

(S1 sample)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(22.90, 23.00)	(30.50, 32.75)	(62.50, 65.50)	(94.00, 96.00)	(140.9,141.2)
CI=0.25	(22.26, 23.00)	(30.50, 33.50)	(60.00, 68.00)	(93.00, 98.00)	(124.8,141.9)
CI=0.5	(22.00, 23.00)	(30.50, 34.50)	(57.00, 70.50)	(92.25, 100.75)	(123.9,146.9)
CI=0.75	(21.80, 24.00)	(25.50, 37.50	(56.00, 74.00)	(91.00, 106.00)	(112.9,153.0)
CI=0.95	(20.00, 25.00)	(22.50, 39.50)	(54.00, 76.00)	(84.00, 111.00)	(86.0, 295.0)

Figure 7-28: Table of the basic bootstrapped runout (S1) values of the sample.

(S1 model)	p=0.1	p=0.25	p=0.5	p=0.75	p=0.95
CI=0.1	(23.15, 23.61)	(33.38, 35.75)	(59.91, 62.20)	(93.25, 94.95)	(148.8,149.7)
CI=0.25	(22.92, 24.81)	(30.81, 35.80)	(58.32, 64.49)	(92.18, 96.63)	(145.7,159.6)
CI=0.5	(22.49, 25.40)	(29.90, 38.98)	(52.98, 66.72)	(89.99, 99.47)	(139.0,163.4)
CI=0.75	(21.46, 26.21)	(28.08, 39.67)	(48.97, 68.73)	(86.49,101.35)	(137.9,167.4)
CI=0.95	(20.20, 27.32)	(24.63, 43.30)	(47.24, 70.11)	(83.17,108.87)	(83.0, 291.1)

Figure 7-29: Table of the basic bootstrapped runout (S1) values of the multiple regression model.

7.6 Threshold values with respect to the γ parameter and height

The different potentials of the mountainside were grouped according to the different height levels and plotted with respect to runout distance beyond the talus toe (S1) and the γ angle. The results had a large scatter and revealed no particular threshold level for the isolated height of the talus (H2) (fig.7-30). The same scattering was seen with respect to the height of the cliff (H3) (fig.7-32), although here the correlations for different levels were slightly better than for H2. The problem of scattering meant that the longest runouts contained comparatively smaller values of H2 and H3. As an example of this, the three longest runouts had H2 values of 59, 50 and 29 meters respectively.

Even though the spread was significant, a lower bound for H2 with respect to the γ values was set, neglecting the six highest H2 values (fig.7.31). This was chosen as there was a definite break between these six and the rest of the distribution. The line is in effect a lower boundary for 95 per cent of the H2 values and is equal to:

(H2<200m) =11.56γ + 96.45 m.

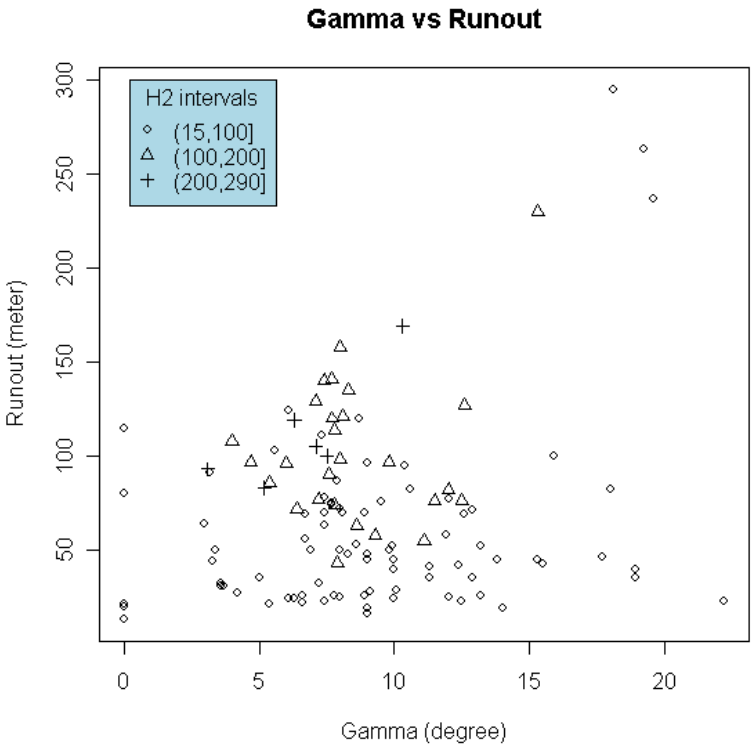


Figure 7-30: Figure displaying Gamma vs. Runout with respect to the height of the talus (H2).

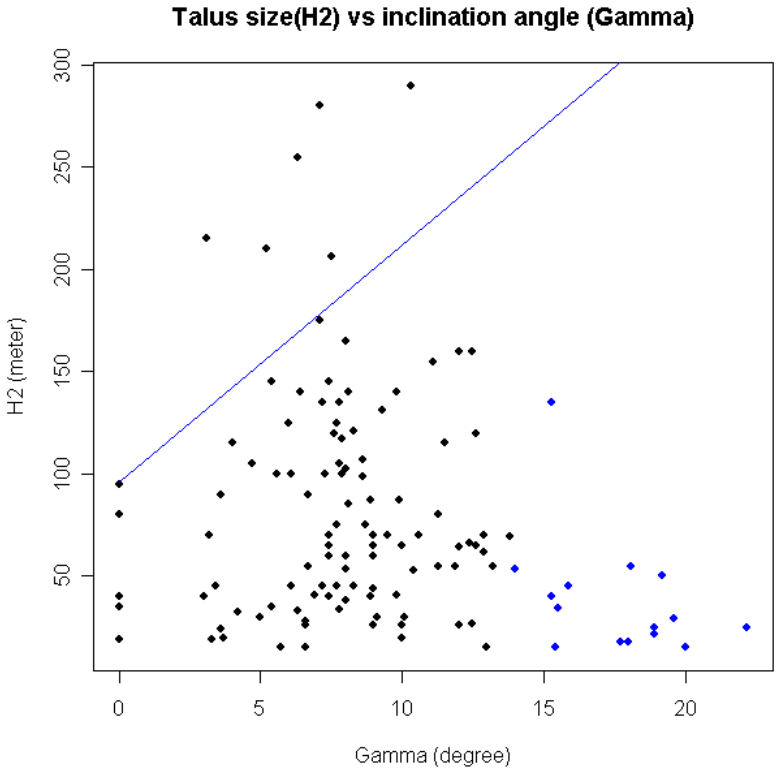


Figure 7-31: γ vs. H2 with γ≥14° in blue. The blue line represent (H2<200m)=11.56γ + 96.45m.

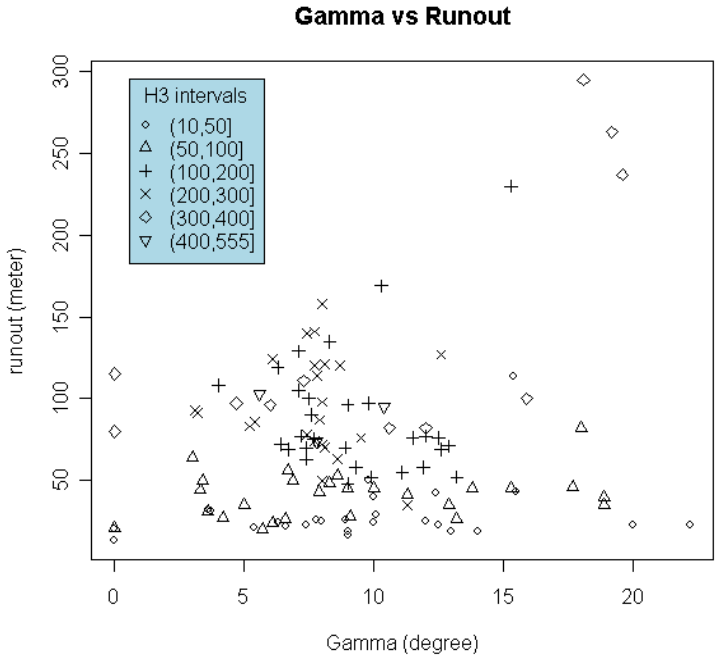


Figure 7-32: The γ parameter vs. runout with respect to the height of the cliff (H3)

Because of the large spread in the relation between the runout and the two potentials, the total potential (htot), consisting of all three heights (H1+H2+H3), was utilized to estimate upper and lower threshold values for the runout distance beyond the talus (S1). For some of the limit lines the point variables were close together. In effect the uncertainty is higher for these lines because they have a relatively small difference in γ angles. Also notable is the fact that some of the threshold lines are based on a small amount of observations. This is particularly true for the highest potential interval between 500 and 625 meters, which only consists of five observations. Furthermore, the observations with the greatest potential were all within the boundary lines for 400 m ≥ htot ≥ 300 m. Thus, the limits for the highest potentials 625≥htot≥500 is not represented in the plot.

If we were to think of the talus of the mountainside as a significant retardation area of the boulder, we would be inclined to believe that the height of the cliff (H3) would be a deciding factor for the runout of the boulder. Thus the analysis confirms an earlier report by Domaas, which assume that the retardation process in the talus is only moderate for the longer rockfall ranges (Domaas 1985). Furthermore, the plot show that the three longest runouts all come from the interval containing the second highest cliffs, while the fourth longest has a much lower cliff height. However, the analysis as a whole does not show any clear relation between the runout range and the height of the cliff (H3).

To further investigate the relation between retardation and height of the cliff (H3), a 3-dimensional scatter plot was made (Appendix J). The plot clearly revealed that the three longest runouts (observation 120, 121 and 122) has comparatively small S2-values, thus the retardation caused by the friction against the scree will be over a relatively shorter distance for these runouts.

However, this is far from a general effect as observation 115 and 116 has much larger taluses. The plot of γ vs. runout beyond the talus produced the following estimated threshold values:

For interval 500≥htot>400:

S1 = 9.977γ + 115 (lower boundary represented by full green line in fig.54)
S1 = 12.797γ - 14.71 (upper boundary represented by stapled green line in fig.54)

For interval 400≥htot>300:

S1= 11.479γ + 54.90 (lower boundary represented by full blue line in fig.54)
S1=14.108 γ - 124.2 (upper boundary represented by stapled blue line in fig.54)

For interval 300≥htot>200:

S1=6.391γ + 82.90 (lower boundary represented by full red line in fig.54)
S1=1.694γ + 28.65 (upper boundary represented by stapled red line in fig.54)

For interval 200≥htot>50:

S1= 4.0γ + 52.61 (lower boundary represented by full grey line in fig.54)

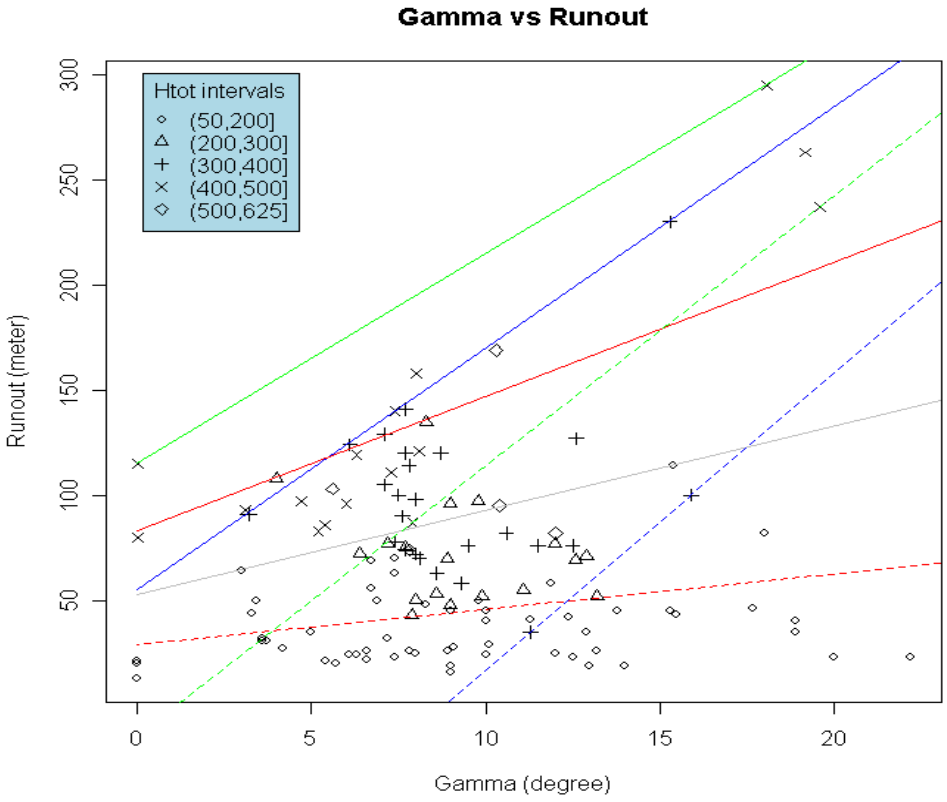


Figure 7-33: Plot with Gamma vs. runout with different total heights (htot).

8. Validation and general remarks

For this thesis all observations have been utilized to develop in all five different model types. As a consequence of employing all observations we do not have any independent test dataset based on the remaining points to validate our results. This weakness has also been commented on by Keylock in his article using the same basis (Keylock 1999). However, an advantage of applying the full dataset is that we get an insight into which observations are most influential in our regression setting. In this respect, some of the outliers are beyond the criteria of neglect. This could be a natural effect due to the general simplification of complex slopes, or because of indistinguishable features in the terrain. Even so, there is no reason why these observations should be significantly erroneous. Thus, for the sake of conservation all observations that produced runouts beyond the talus were used in the analysis.

The basic bootstrap method revealed definite signs of instability, particularly for higher percentile values with high confidence intervals. This is believed to be related to the fact that there are fewer observations to sample from at higher percentile levels. Thus, the bootstrapped intervals for distributional extreme values will be prone to vary more relative to lower percentile values. Additionally, in the light that bootstrapping represents a stochastic form of statistic, the method of comparing the modeled values with that of the sample must be seen as estimation between the models.

In order to get a better overview of the models performance, a table (fig.8-1) displaying each model's result with respect to the original sample ((S1+S2) and S1) was made. For the sake of completeness a bootstrap was also conducted on the energy loss model derived by separate friction coefficients. For the multiple model the bootstrapped values are based on the S1 values.

Model type	***(R^2)***	***Regression coefficient***	***Standard model error (m)***	***(S1+S2) p=0.1 CI=0.95***	***(S1+S2) p=0.5 CI=0.95***	***(S1+S2) p=0.95 CI=0.95***
Statham (avg.)	0.92	0.89	35.1	(68.44, 86.92)	(154.0,241.0)	(309.8,469.5)
Energy loss (avg.)	0.86	1.03	56.0	(68.16, 94.58)	(162.7,221.9)	(287.0,569.3)
Energy loss	0.99	1.04	16.3	(67.27, 85.84)	(167.6,241.2)	(372.0,521.4)
α - δ	0.96	1.11	32.5	(47.23, 77.41)	(163.4,225.3)	(366.7,619.8)
β – Ψ	0.88	1.04	50.8	(65.62, 84.98)	(155.8,249.5)	(369.4,536.1)
Multiple	0.97	-	9.4	(20.20, 27.32)	(47.24,70.11)	(83.0,291.1)

Figure 8-1: Table of all model results. The p signify the quantile level, while the confidence level is constant at 0.95.

From the table we see that the multiple regression model yields the most accurate results with a residual error of only 9.4 meters and a squared correlation value of 0.97 for the runout beyond the talus (S1). As stated earlier the model effectively manages to capture most of the distributional extremes, as well as the observations at lower percentile levels.
The tested dynamic models showed some variance according to the used method.

The utilization of the Statham model with an average friction coefficient of 0.68 yielded a squared correlation value of 0.92. However, as noted by Keylock the spread is relatively large and it does not hold a significant advantage over the empirically based models. The use of the simplified energy model with separate friction coefficients derived by the tangent of the shadow angle (α) yielded encouraging results considering its simplicity. Even so, the reader should be well aware that even though the model may yield adequate results, the velocity calculated at each level may deviate to that of reality.

8.2 Recommendations for future development

The restitution, shape and volume of the boulder have not been extensively researched or commented on in this study. These factors are of vital importance in rigorous models and should therefore be prioritized for future investigations.

A rough approximation of physical attributes as restitution, velocity and bounce heights in the different slopes may be found by utilizing a stochastic simulation program. In this respect the simulation program Rocfall 3.0 was briefly tested on eight of the observations representing different percentile levels. The analysis was conducted assuming three types of material properties with respect to the restitution at different parts of the mountainside. In this respect the author chose the default restitution values of clean bedrock, talus and vegetated talus from mountain top to bottom respectively. Furthermore, the exact coordinates of the slopes were plotted for each slope. The results, carried out with 50 rocks and a mass of 40 kg, revealed as expected a significant variance. However, when founding the tangential restitution on the mass and using a friction angle derived from this, around 10 or fewer rocks reached the end of the mountainside. In the end it is difficult to say if this is realistic or not, even so the relations are interesting enough to encourage further investigations.

9. Conclusion

A total of five different model types for prediction of rockfall runout range have been tested in this thesis. Of these, two are dynamical and three are based on empirically based angles from a mountainside. The results indicate that the simplified dynamic models have a relatively large variance when based on an average dynamic friction angle ($\phi_{\mu}^{'}$). In this respect, the Statham model is clearly less sensitive than the energy loss model. However, when utilizing separate friction values a relatively high level of accuracy was achieved with the latter model. This indicate that a simplification to that of a dry Mohr-Coulomb mass on the boulders principal movement part, that is the talus and the terrain beyond, may yield an adequate first estimation of runout distance for rockfalls. Furthermore, a tangential energy loss of about 70 -75 per cent of the kinetic energy during impact with the bottom of the cliff shows encouraging results when used in context with the simple energy loss model described in chapter 6.2. This suggests that the shadow angle is an important parameter with respect to rockfall runout range. As a consequence of the model performance, the relation between runout and the shadow angle should be investigated further. Even so, the result of the simplified models utilizing an average dynamic friction coefficient does not produce significantly better results than the empirical models.

Among the empirically derived models, the multiple regression model yields the most accurate results with a squared correlation value of 0.97 when compared with the original sample basis. The relations in the model indicate that there is two important two way interactions; between the talus angle (ψ) and the total height (htot), and between the fahrböschung (β) and the total height (htot). Furthermore, the thesis confirms that the inclination beyond the talus as signified by the γ angle is of vital importance for the runout beyond the talus. In this respect, a limit value of 14^{o} was found among the selected parameters by use of a weighted maximum likelihood method in a regression tree.

In the models derived from the simple linear regression, the most promising result was found between the empirical β-ψ relation. The model based on the correlation between the parameters produced lengths which were somewhat underpredicted for quantiles between the 10^{th} and 50^{th} percentile, while the longest distances around the 95^{th} percentile, were somewhat underpredicted from that of the original sample base. The other tested simple linear model was based on the α-δ relation. The model was based on a significantly lower correlation value than the β-ψ relation. However, as the largest spread was found for the observations of relatively high exceedance probabilities between 1 and 0.5, the correlation would rise significantly with the addition of the S2 length. As a consequence, the model is a better approximator than the β-ψ model for slopes with the potential of relatively longer runout distances beyond the talus (S1). The statistical results in this thesis prove that the empirical relations yield a reasonable first estimate of rockfall runout range. In effect this means that a combination of the two simple linear models could be used to give a reasonable first estimation of runout range. Ultimately, the selection of a preferred model is dependent on the amount of data available from a given region and the acquired level of safety conservation.

Acknowledgements

During the writing of this thesis many have stood by to help me out. In this respect I would like to thank my main supervisor, chief senior engineer Ulrik Domaas at the Norwegian Geotechnical Institute (NGI) for his general support, and for assisting me with insightful comments and suggestions. He has always been enthusiastic and I appreciate that he has been able to discuss the thesis with me during a busy schedule.

Professor Kaare Høeg (UIO/NGI) has been the co-supervisor for this thesis. He has been a mentor and support for me in the field of engineering geology and geohazards throughout my masters degree. His expertise, overall support and final inputs were very much appreciated. Also highly appreciated was the advice and support from ph.D. student Martin Morawietz considering the possibilites within the R software.

I have had many interesting discussions with the colleagues at my study room. In this respect I would have to thank Egil Syre for his general thoughts on my thesis and contribution with respect to computer formatting. Last, but not least I would like to give my warmest thanks to my family for their continuous support and upbringing.

References

Alejano, L.R.P., B.P Bastante, F.G Alonso, E. and Stockhausen, H.W. . 2007. Slope geometry design as a means for controlling rockfalls in quarries. *International Journal of rock and Mining Sciences* 44, 903-921.

Azzoni, A. 1993. *Methods for predicting rockfalls*. London. Imperial college of Science, Technology and Medicine (University of London).

Berger, F.a.D., L.K.A. 2007. Principles of the tool Rockfor.net for quantifying the rockfall hazard below a protection forest. *Schweiz Z Fortswes* 158, 157-165.

Bozzolo, D.a.P., R. 1986. Simulation of Rock Falls down a valley side. *Acta Mechanica* 63, 113-130.

Brown, A.G. 1994. *The internal structure of talus deposits*. Report, Institute for Natural Geography, UIO, Oslo. 156.

Crawley, M.J. 2005. *Statistics: An introduction using R*. Edited by Sons, J.W., Chichester: John Wiley & Sons,Ltd.

Davis, J.C. 2002. *Statistics and data anlysis in geology*. Edited by Gerber, M.a.B., Brienna. 3 ed, New York: John Wiley & Sons. 638 pp.

Domaas, U. 1985. *Rekkevidden av steinsprang*. Oslo. NGI, 58500-1. 17 pp.

Domaas, U. 1994. *Geometrical methods of calculating rockfall range* Oslo. NGI, 585910-1. 21 pp.

Dorren, L.K.A. 2003. A review of rockfall mechanics and modelling approaches. *Progress in Physical Geography* 27, 69-87.

Dorren, L.K.A. and Seijmonsbergen, A.C. 2003. Comparison of three GIS-based models for predicting rockfall runout zones at a regional scale. *Geomorphology* 56, 49-64.

Efron, B.a.T., Robert J. 1993. *An Introduction to the bootstrap*. Monographs on statistics and applied probability, New York: Chapman & Hall. 436 pp.

Fitzmaurice, G. 2000. The meaning and interpretation of interaction. *Nutrition* 16, 313-314.

Freund, R.J.a.W., William J. 1998. *Regression analysis: Statistical modeling of a response variable*, San Diego: Academic Press.

Hastie, T.J.a.T., R.J. . 1990. *Generalized Additive Models* New York: Chapman and Hall. 335 pp.

Heidenreich, B. 2004. *Small and half scale experimental studies of rockfall impacts on sandy slopes* Ph.d Thesis, Rock Mechanics laboratory (LMR) of the Swiss Federal Institute of Technology of Lausanne, ELP, Lausanne. 226.

Hungr, O.a.E., S.G. 1989. *Engineeering Aspects of Rockfall Hazards in Canada*. Vancouver, B.C. 102 pp.

Keylock, C.a.D., Ulrik. 1999. Evaluation of topographic models of rockfall travel distance for use in hazard application. *Arctic, Antarctic and Alpine research* 31.

Krumbein, W.C.a.G., Franklin A. 1965. *An introduction to statistical models in geology*. Edited by Shrock, R.R.: McGraw-Hill Book Company. 475 pp.

Larsen, J.O. 1993. *Rockfall simulation models and their practical application*. Report, University of Colorado Boulder.

Moore, D.S.M., George P. 2003. *"Introduction to the practice of statistics"*. Edited by Farace, P. 4 ed: Julet, Michelle W.H. Freeman and Company. 817 pp.

Petje, U.e.a. 2005. Computer simulation of stone falls and rockfalls *Acta geographica Slovenia* 45, 102-103.

Pfeiffer, T.J.a.H., J.D. 1990. Rockfall hazard analysis using the Colorado rockfall simulation program. *Transportation research record 1288:geotechnical engineering*, 118-119.

Ripley, B.*The tree package*. http://cran.r-project.org/web/packages/tree/tree.pdf

Statham, I. 1972. Scree slope development under conditions of surface particle movement. *The relationship between scree slope morphology and the dynamics of small-scale rockfalls*, 41-52.

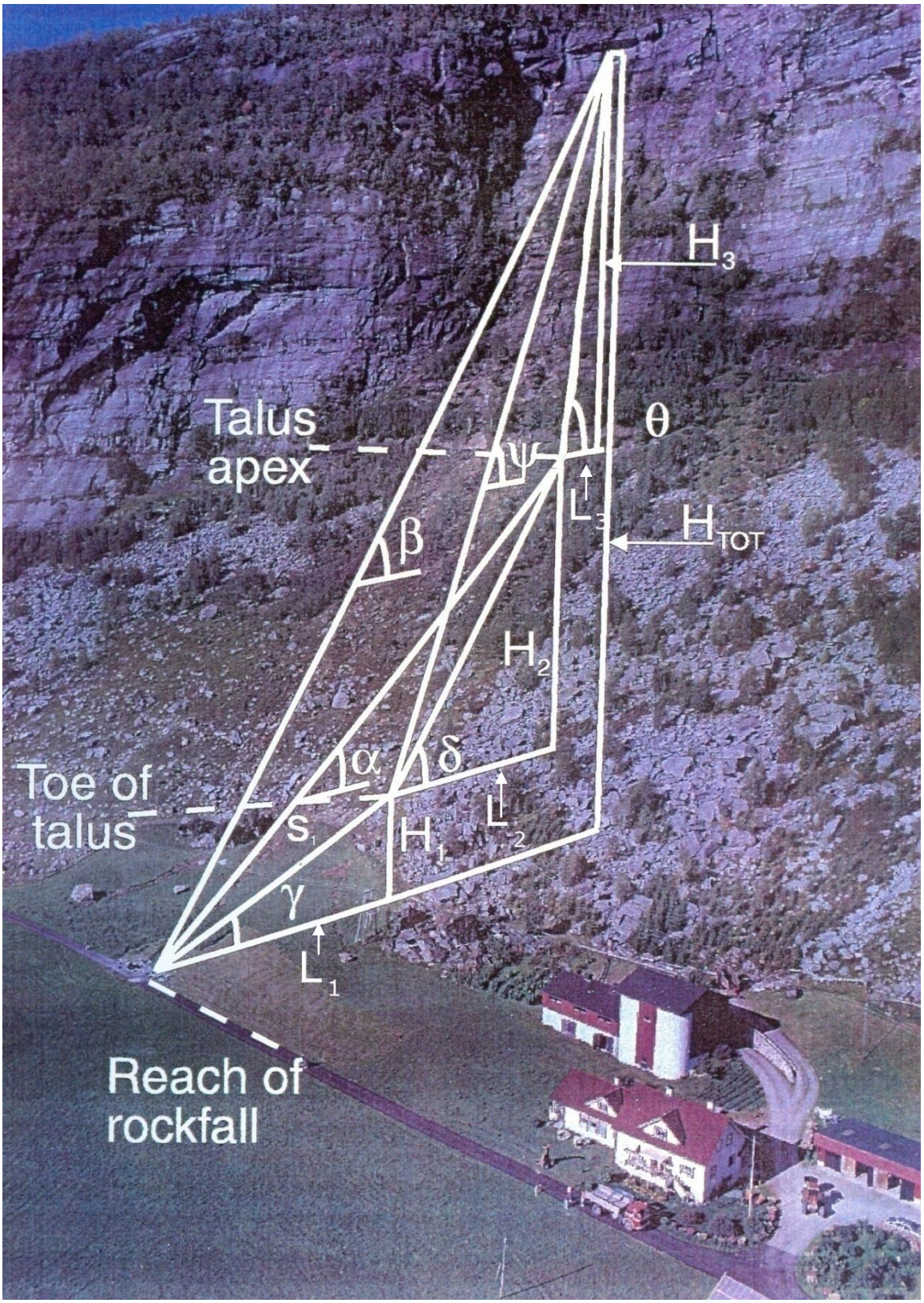
Statham, I.a.Kirkby., M.J. 1975. Surface stone movement and scree formation. *Journal of Geology* 83, 349-362.

Teknomo, K. 2006. Bootstrap sampling tutorial. http://www.people.revoledu.com/kardi/tutorial/bootstrap

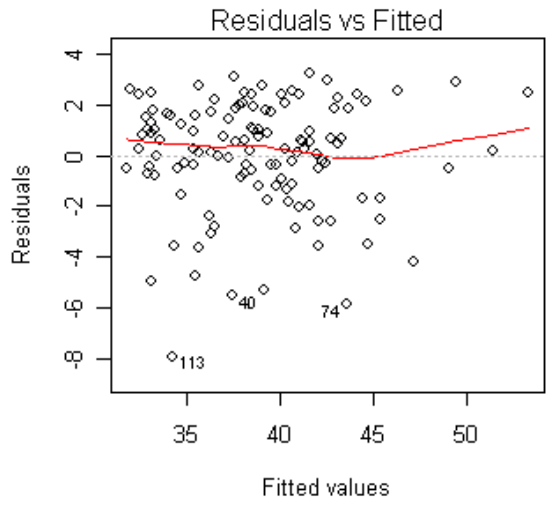
List of figures

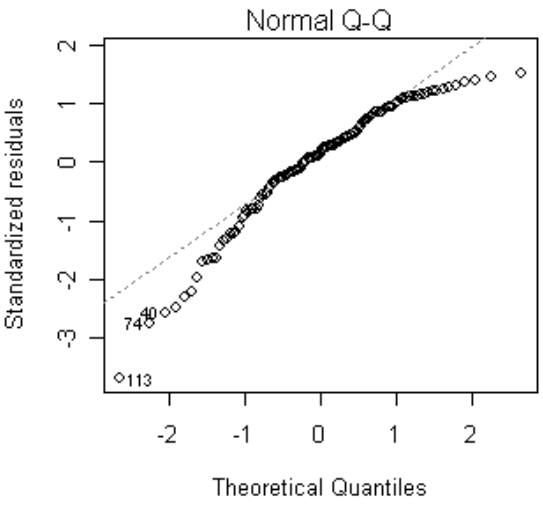
Appendix A-J

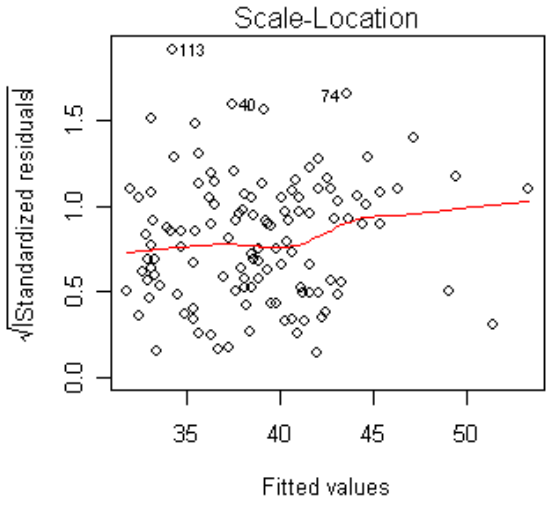
Appendix A: Photography of mountainside with associated terrain parameters utilized in empirical models

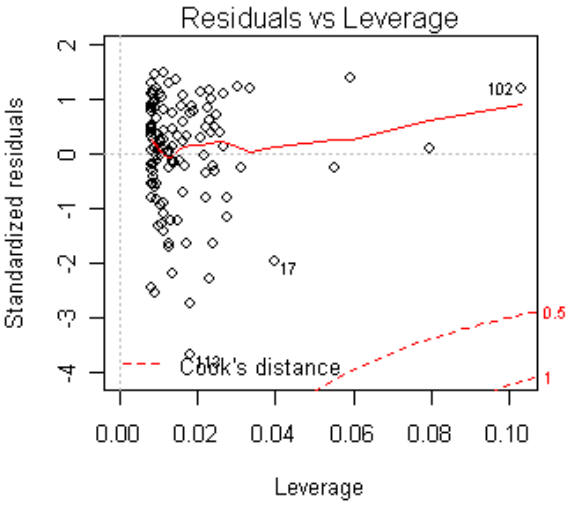


Appendix B: Diagnostic plots for β-ψ relation

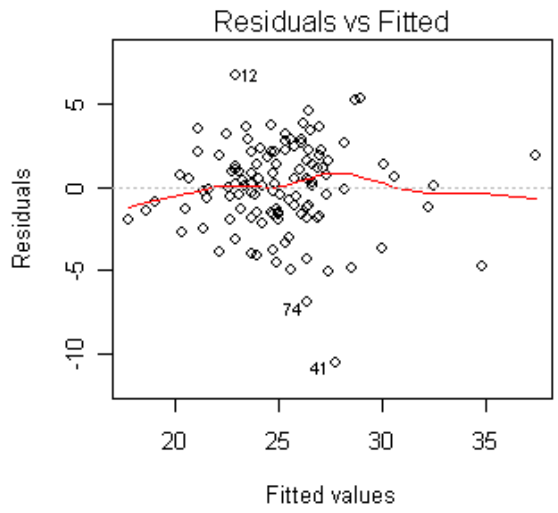


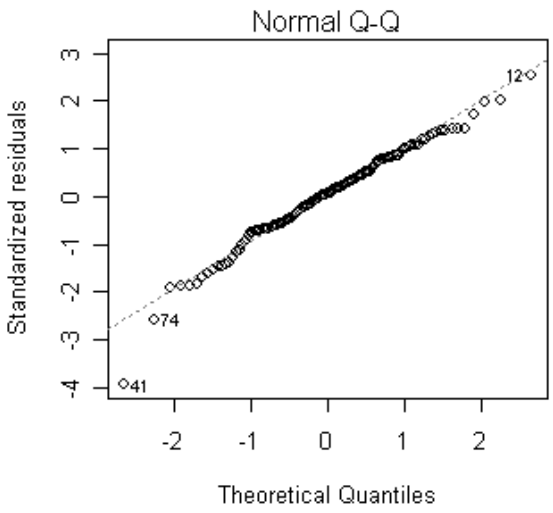


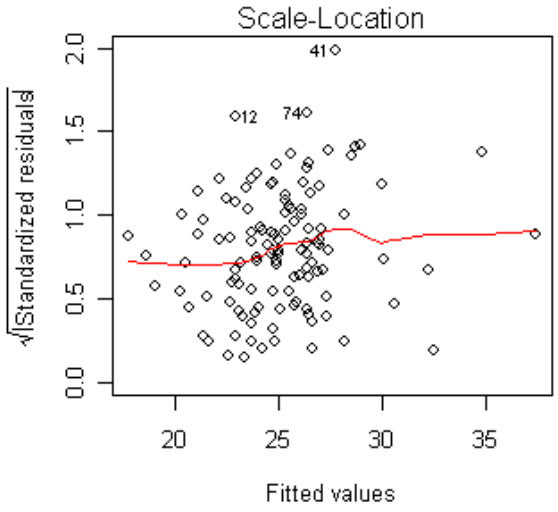


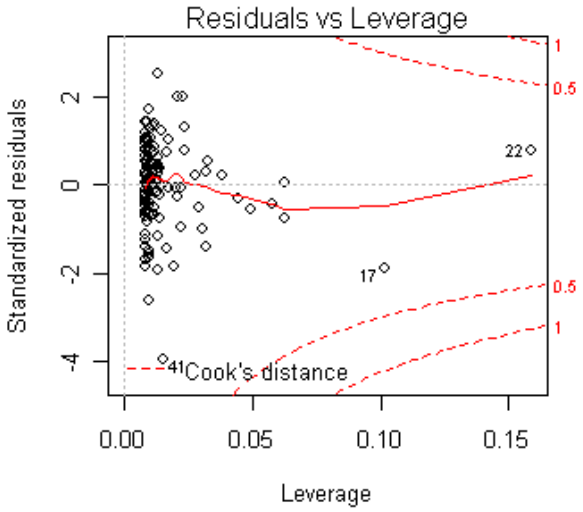


Appendix C: Diagnostic plots for α-δ relation

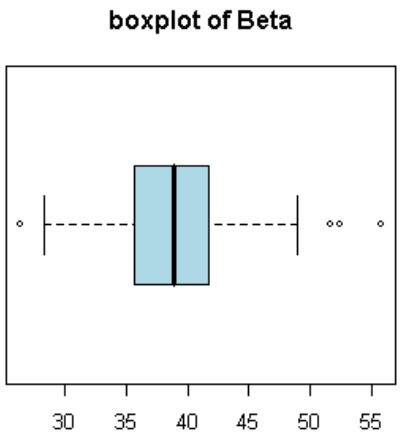


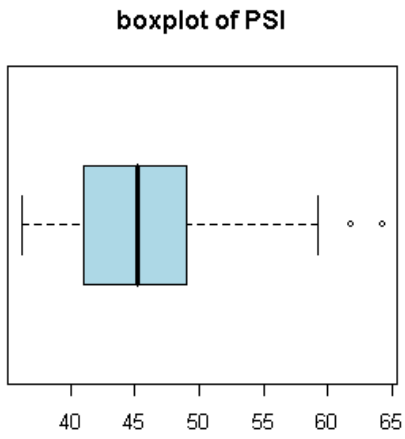


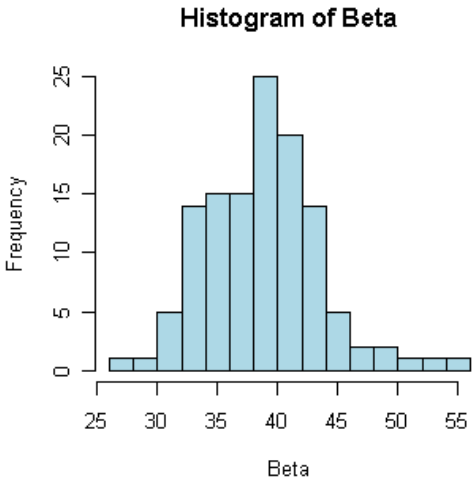


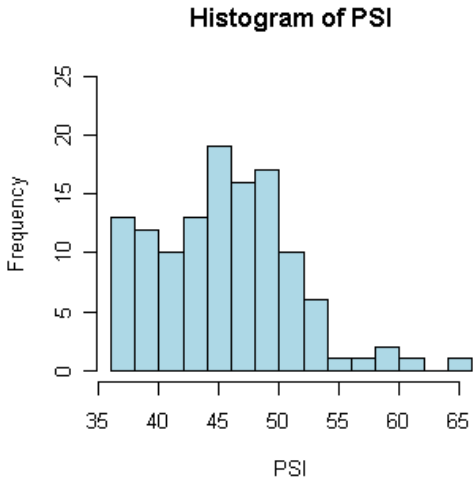


Appendix D: Graphical display of β and ψ distributions

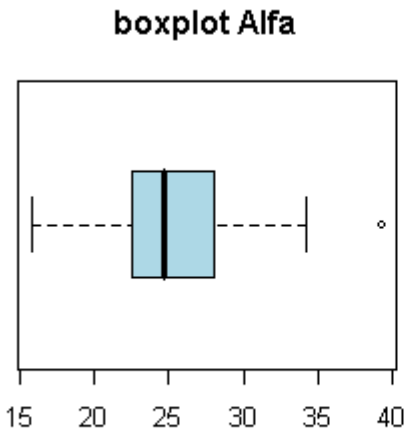


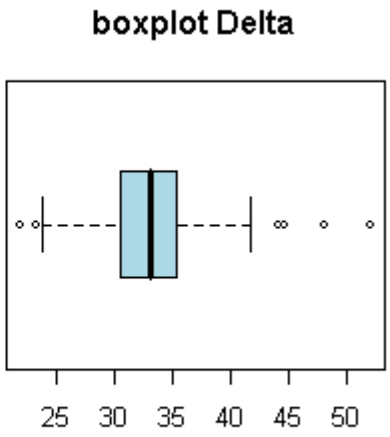


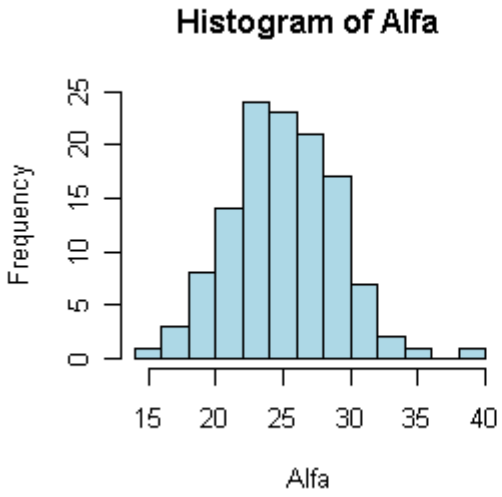


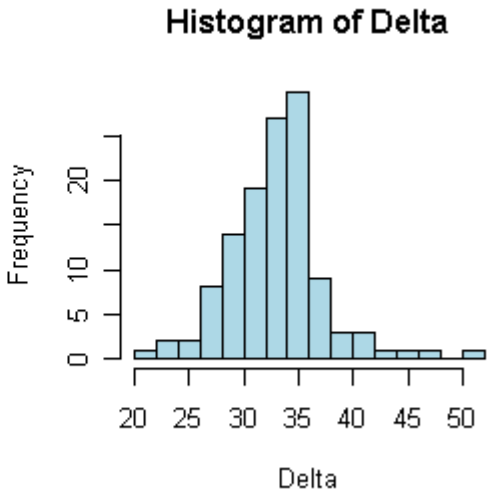


Appendix E: Graphical display of α and δ distributions









Appendix F: Initial multiple model and related statistics

Model Call:
lm(formula = Runout ~ Htot + I(Htot^2) + Beta + I(Beta^2) + Alfa + I(Alfa^2) + PSI + I(PSI^2) + Gamma + I(Gamma^2) + Delta + I(Delta^2))

- **Least significant terms are marked in red**

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.445e+02	1.068e+02	-2.289	0.023991	*
Htot	5.038e-01	5.748e-02	8.764	2.78e-14	***
I(Htot^2)	-3.667e-04	1.008e-04	-3.639	0.000420	***
Beta	-1.480e+01	7.742e+00	-1.911	0.058610	.
I(Beta^2)	5.750e-02	1.001e-01	0.574	0.566945	
Alfa	1.586e+00	7.094e+00	0.224	0.823553	
I(Alfa^2)	-4.343e-02	1.367e-01	-0.318	0.751241	
PSI	2.110e+01	8.199e+00	2.573	0.011412	*
I(PSI^2)	-1.423e-01	8.898e-02	-1.599	0.112743	
Gamma	-2.370e+00	1.637e+00	-1.448	0.150577	
I(Gamma^2)	2.600e-01	7.454e-02	3.489	0.000702	***
Delta	1.795e+00	5.885e+00	0.305	0.760891	
I(Delta^2)	-2.440e-02	8.458e-02	-0.289	0.773506	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.58 on 109 degrees of freedom

Multiple R-Squared: 0.8446, Adjusted R-squared: 0.8275

F-statistic: 49.38 on 12 and 109 DF, p-value: < 2.2e-16

Appendix G: Final multiple model and related statistics

Model call:
lm(formula = runout ~ htot + PSI + I(Gamma^2) + htot:Beta + htot:PSI + htot:PSI:Gamma)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.467e+01	1.556e+01	4.156	6.26e-05 ***
htot	3.118e-01	5.109e-02	6.104	1.44e-08 ***
PSI	-1.490e+00	3.430e-01	-4.345	3.02e-05 ***
I(Gamma^2)	5.102e-02	1.443e-02	3.537	0.000585 ***
htot:Beta	-5.773e-02	1.927e-03	-29.955	< 2e-16 ***
htot:PSI	4.856e-02	1.829e-03	26.553	< 2e-16 ***
htot:PSI:Gamma	8.866e-05	2.542e-05	3.488	0.000690 ***

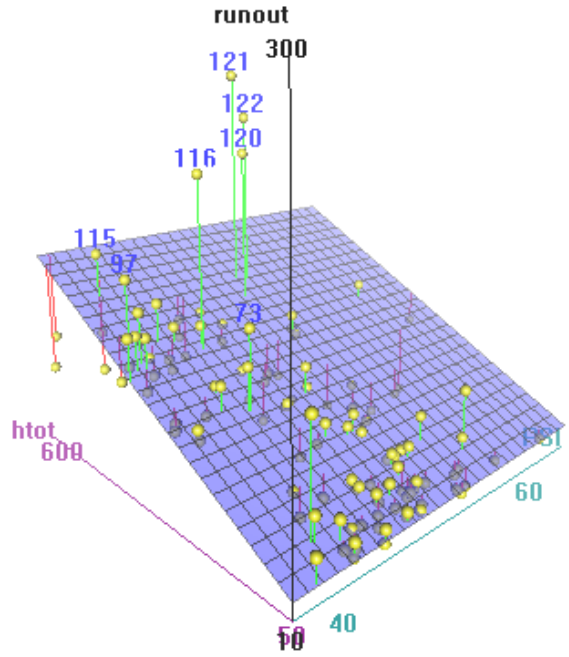
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

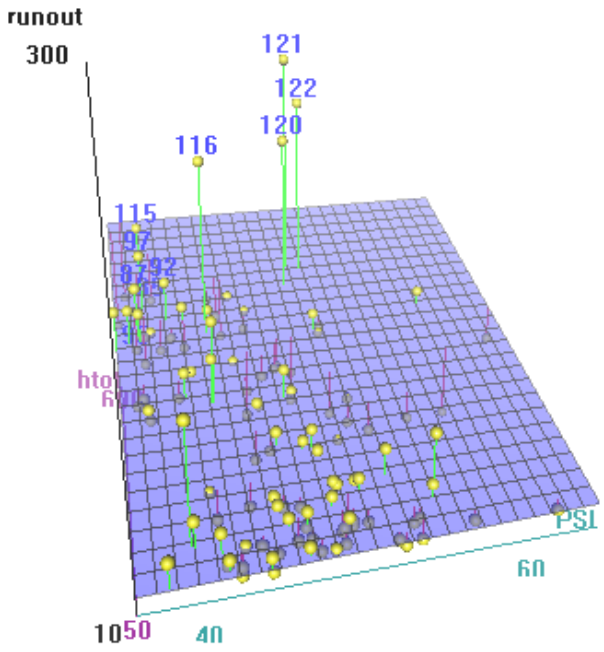
Residual standard error: 9.267 on 115 degrees of freedom

Multiple R-Squared: 0.9668, Adjusted R-squared: 0.965

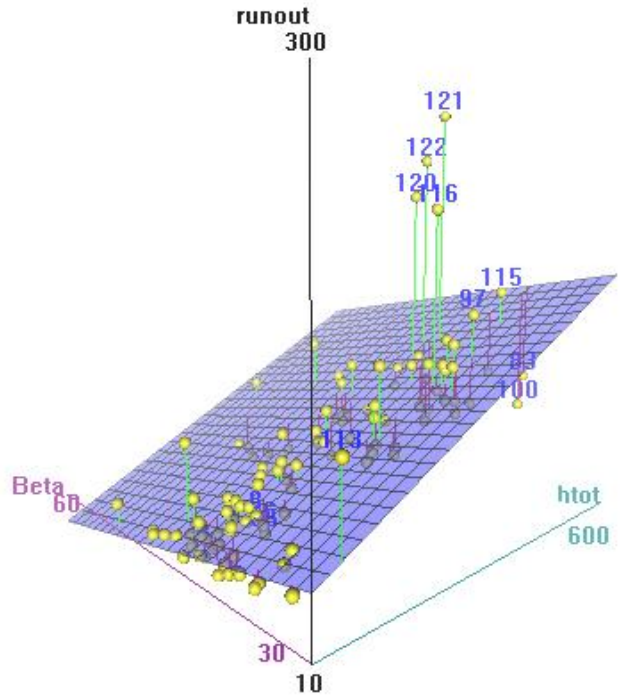
F-statistic: 557.4 on 6 and 115 DF, p-value: < 2.2e-16

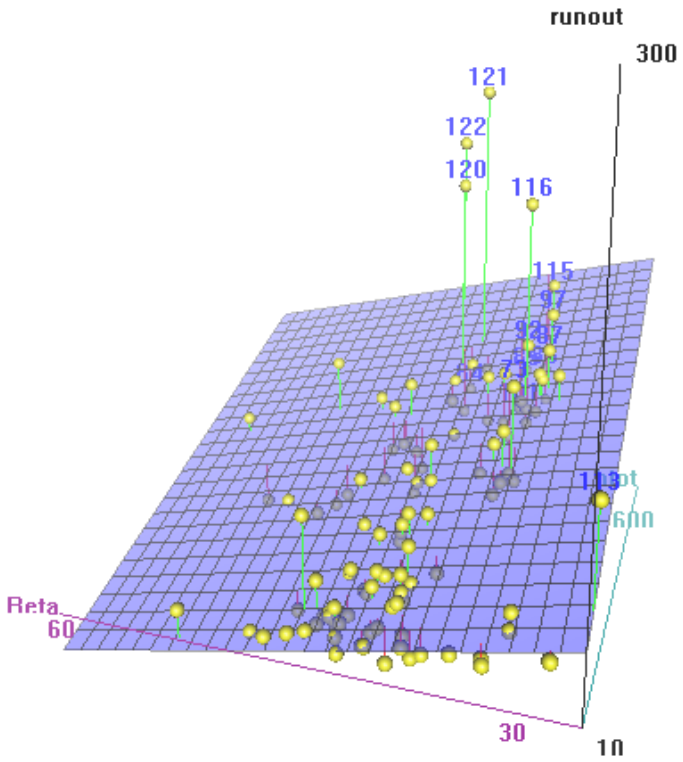
Appendix H: 3D plot of S1 as a response of Htot and ψ





Appendix I: 3D plot of S1 as a response of β and Htot





Appendix J: 3D plot of S1 as a response of H3 and S2